The roles of recombination in trinucleotide repeat instability in *E. coli*

Stuart A. Warner

Submitted for a Molecular Biology PhD from The University of Edinburgh, 2002.

Declaration

I declare that I am the sole author of this thesis, that the work presented in it is all my own, and that it has not been submitted previously for any other degree or professional qualification.

Declaration 2
Table of contents 3
List of figures 7
List of tables 9
List of abbreviations, units and prefixes 10
Acknowledgements 13
Abstract 14

Chapter 1: **Introduction** 16
Triplet Repeat Expansion Diseases (TREDs) 16
The 9 expressed polyglutamine diseases 17
Inherited disorders associated with unexpressed TR tracts 20
Folding properties of TRs 24
The study of palindromic sequences has aided understanding of DNA hairpins and TR tract pseudohairpins 28
Hairpin-mediated instability and inviability 30
cruciform extrusion 31
Microsatellite Instability (MIN) 36
Recombination in *E. coli* 38
The RecB pathway of double strand break repair 39
The RecF pathway of single strand gap repair 42
The *recF*⁻ *recO*⁻ *recR*⁻ epistasis group 42
RecF 44
RecR 44
RecO 45
RecJ 45
RecQ. 46
RecN 47
SbcCD nuclease. 48
RecA is the agent of synapsis in Homologous Recombination. 49
Postsynapsis 51
RuvABC 51
RecG 54
The loosening distinction between recombination and replication. 56
D-loop stabilisation converts a common recombination intermediate into a substrate for initiation of DNA replication 57
Replication initiation by PriA 57
Replication is needed for recombination 59
Inducible Stable DNA Replication (iSDR) is initiated at recombination intermediates during SOS de-repression 62
Constitutive stable DNA replication (cSDR) occurs at R-loops in *recG*⁻ and *rnhA*⁻ mutants 63
DNA replication in pUC18-derived plasmids. 64
The SOS response 66

Chapter 2: **Materials and Methods** 70
Materials 70
Strains of Bacteria 70
Plasmids 74

Synthetic Oligonucleotides 75
Media 77
Media Additives 78
General Stock Solutions 80
Qiagen Solutions for extraction of plasmid DNA from *E. coli* 82
Solutions for Agarose and Polyacrylamide Gel Electrophoresis 82
Restriction Endonucleases and Incubation Buffers 83
Materials for Polymerase Chain Reactions 84
Materials for end-labelling of DNA 85
Methods 86
Bacterial Culture Methods 86
Storage of Bacteria 86
Growth of Overnight Cultures 86
Mitomycin C incubations and safe disposal 86
β-galactosidase assay 86
Transformation of Bacterial cells 87
Preparation and Storage of $CaCl_2$-Competent *E. coli* cells 87
Transformation of $CaCl_2$-Competent *E. coli* Cells 87
Preparation of Electrocompetent *E. coli*. 88
Transformation of electrocompetent *E. coli* 88
Transfer of mutations between *E. coli* strains by P1 transduction 88
Preparation of P1 Plate Lysate 89
Storage of bacteriophage 89
Measuring the titre of a P1 lysate 89
P1 Transduction 90
Bacteriophage Lambda Protocols 90
Harvesting bacteriophage Lambda from plate lysates 90
Lambda phage spot test 91
Methods of DNA Purification 91
Phenol/Chloroform purification of DNA 91
DNA Precipitation 91
Extraction of Plasmid DNA from *E. coli* using Qiagen mini-prep and Maxi-prep kits 92
Crude *E. coli* genomic DNA extraction 93
Purification of PCR Products 94
Extraction of DNA Fragments from Agarose Gels 94
DNA quantification 95
Measurement of the Concentration and purity of DNA 95
Estimation of the relative proportion of two or more DNA bands in an agarose gel 96
Molecular Manipulation of DNA 96
DNA Restriction Digests 96
The Polymerase Chain Reaction (PCR) 96
Radioactive End-Labelling of Double-Stranded DNA Fragments 98
Autoradiography 99
Seperation of DNA species and fragments using gel electrophoresis 99
Agarose Gel Electrophoresis 99
Native Polyacrylamide Gel Electrophoresis 100
Sterilisation of solutions 101
Autoclaving 101

Sterile filtering 101

Chapter 3: **Creation of strains containing complete knockouts of *recQ*, *recF*, and *yraN*.** 102
Rationale 102
Genes chosen for complete removal of activity 102
RecQ and its homologues 102
WRN 103
BLM 104
yeast Sgs1 104
Fission yeast Rad12/Rqh1 105
RecF 106
YraN 107
Engineering of the *E. coli* chromosome using activated double strand break repair pathways. 110
Production of transforming DNA constructs using PCR 114
Preparation of DL49 and activated DL1218 electrocompetent cells. 115
Selection of mutant colonies 115
Results and discussion of PCR cloning experiments 117
YraN is not a novel resolvase 119
Summary 122

Chapter 4: **Multimerisation of plasmids containing mouse $(CAG/CTG)_{43}$ trinucleotide repeat tracts** 123
Aims 123
Introduction to plasmid multimerisation 123
Recombinational models of plasmid dimerisation 127
RecA-independent mispairing between direct repeats 128
The cruciform/dumbell model for inverted dimer production 130
Methods 131
Results 134
General levels of pUC18 dimerisation in each strain 137
TR tract orientation dependence on plasmid dimerisation 140
Discussion 141
Summary 146

Chapter 5: **Do CXG trinucleotide repeat tracts induce de-repression of the SOS response?** 148
Aims 148
Causes of SOS de-repression 149
Experiments using the *sfiA:lac* SOS reporter system 150
Measurement of β-galactosidase activity 151
Results and discussion of the β-galactosidase assay experiments 157
Experiments using the *λ-gal* SOS reporter system 159
Methods for the *λ-gal* SOS reporter assay 161
Results of the *λ-gal* SOS reporter assay in MT101. 162
The *λ-gal* SOS reporter assay in *recD*⁻ cells. 164
Results to *λ-gal* SOS reporter assay in *recD*⁻ cells (DL1555) 165
Unified discussion of SOS reporter assays 167

Conclusions 172

Chapter 6: **Competition between plasmids with and without trinucleotide repeat tracts.** 173
Aims 173
Methods 176
Results 178
Sources of variation 183
The sign test 184
Total plasmid size was not a factor in determining the outcome of competition. 185
Loss of mouse CAG/CTG_{43} did not result from deletion of the TR tract. 185
The effect of triplet sequence on competition. 186
TR orientation has little effect on competition 186
The presence of mouse flanking sequences is not an important factor in competition 187
Reduced competition was observed in the *umuDC*⁻ background. 187
Discussion 188
The relative loss of mouse CAG/CTG_{43} 188
Possible loss via inhibited replication through long TR tracts 188
Possible loss via endonucleolytic cleavage of hairpin structures in TR tracts 189
Possible loss via the formation of linear multimer plasmids 190
Possible loss of long TR plasmids by delayed recovery from stationary phase 191
The relative gain of short TRs 192
Replication can be initiated from stable R-loops 192
DSBR of *chi*-less plasmids in SOS-induced cells may favour replication (iSDR) rather than plasmid degradation 193
Reduced competition in *umuDC*⁻ cells 196
Summary 199

Chapter 7: **End-labelling experiments to determine trinucleotide repeat tract instability in different DNA replication and repair mutants in *E. coli*** 201
Aims 201
Methods 202
Results 204
General observations 212
Artificial CCG_{24} 213
Artificial CGG_{24} 213
Mouse CAG_{43} 213
Mouse CTG_{43} 213
Discussion 215
General stability observations 215
Orientation effects on TR tract instability 216
Recombination mutants 217
Secondary structure processing 220
SOS induction 221
DNA damage checkpoint activation 222
Mismatch repair 224
Polymerase mutations 225

UV excision repair 230
Conclusions 231

Chapter 8: **Concluding remarks** 234

Appendix1: Mouse flanking sequences. 237

Bibliography 238

List of figures

Figure 1.1: CGG tract expansion in the *FMR1* gene reduces transcription. p21
Figure 1.2: RNA Polymerase entrapment due to triplex structures within the FRDA gene. p23.
Figure 1.3: Non Watson-Crick binding capabilities of thymine. p25
Figure 1.4: One form of hairpin structure predicted to form within a CTG repeat tract. p25
Figure 1.5: Guanine has the ability to form four-base quartets. p26
Figure 1.6: Quadruplex DNA may form in CGG repeat tracts. p26
Figure 1.7: Proposed hairpin cleavage and recombinational repair based on genetic evidence from studies using long palindromes. p29
Figure 1.8: Recombinational repair of processed cruciforms or offset hairpins could lead to changes in tract length. p32.
Figure 1.9: Replication strand-slippage could lead to changes in tract length. p34.
Figure 1.10: The RecBCD pathway of double strand break repair. p41.
Figure 1.11: The RecF pathway of ss gap repair. p43.
Figure 1.12: Stalled fork reversal and processing leads to non-crossover products.p60.

Figure 2.1: Radioactive labelling of EcoRI restriction fragments by incorporation of α^{35}S-dATP by Klenow enzyme. p98.

Figure 3.1: The UPF0102 family. p108.
Figure 3.2: Model of YraN bound to a 4-way DNA (Holliday) Junction. p109.
Figure 3.3: The overall strategy of recombination-mediated chromosome engineering. p113.
Figure 3.4: A mutation in *yraN* does not modify the UV survival of resolvase-deficient *E. coli*. p120.

Figure 4.1: Processing of specific recombination substrates is biased towards crossover or non-crossover products. p125.
Figure 4.2: Crossover and non-crossover products resulting from recombination at broken replication forks. p126.
Figure 4.3: A model for RecA-independent mis-pairing between direct repeats. p129.
Figure 4.4: The cruciform/dumbell model for inverted dimer production. p130.
Figure 4.5: Plasmid CTG_{43} has an elevated rate of dimerisation in DL1368 compared to pUC18 and CAG_{43}. p134.
Figure 4.6: The proportion of supercoiled dimer expressed as a percentage of the total monomeric and dimeric supercoiled plasmid DNA obtained from different backgrounds. p135.

Figure 4.7: Mean arcsin values for strains in which trinucleotide repeat orientation has no significant effect on plasmid multimerisation. p139.
Figure 4.8: Mean arcsin values for strains showing $(CTG)_{43}$-induced plasmid multimerisation. p139.
Figure 4.9: Hairpin cleavage in an unfinished Okazaki fragment. p147.

Figure 5.1: Regulation of the *sfiA:lac* SOS reporter system. p150.
Figure 5.2: Beta-galactosidase activity is elevated in MG1655 in the presence of pUC18-derived plasmids. p152.
Figure 5.3: Beta-galactosidase activity in MG1655 is minimal in the absence of pUC18-derived plasmids. p154.
Figure 5.4: TRs present in plasmids not derived from pUC18 do not increase beta-galactosidase activity in MG1655. p154.
Figure 5.5: Beta-galactosidase expression from the sfiA promoter in MG1655 is proportional to MitomycinC concentration. p155.
Figure 5.6: Regulation of the lambda lytic and lysogenic cycles. p159.
Figure 5.7: Conditions for SOS induction can be measured in DL1437 by assaying for galactokinase activity on MacConkey media. p160.
Figure 5.8: None of the pUC18 derivatives induced cI repressor autocleavage in DL1437. p162.
Figure 5.9: The addition of MMC to DL1437 carrying pUC18 does induce cI repressor autocleavage. p163.
Figure 5.10: None of the pUC18 derivatives induced cI repressor autocleavage in DL1555. p165.
Figure 5.11: The addition of MMC to DL1555 does induce cI repressor autocleavage. p166.

Figure 6.1: Intercell competition. p174.
Figure 6.2: Intracell competition. p174.
Figure 6.3: example agarose gel from competition studies. p178.
Figure 6.4: The proportion of artificial CCG24 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p179.
Figure 6.5: The proportion of artificial CGG24 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p179.
Figure 6.6: The Proportion of CAG43 repeat plasmid in a mixed population with pUC18 vector, before and after a 12 day incubation. p179.
Figure 6.7: The proportion of mouse CTG43 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p179.
Figure 6.8: The proportion of artificial CAG28 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p180.
Figure 6.9: The proportion of artificial CTG28 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p180.
Figure 6.10: The proportion of mouse CAG25 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p180.
Figure 6.11: The proportion of mouse CTG25 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells. p180.
Figure 6.12: The proportion of mouse CAG43 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells. p181.
Figure 6.13: The proportion of mouse CTG43 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells. p181.

Figure 6.14: The proportion of artificial CCG24 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells. p181.
Figure 6.15: The proportion of CGG24 in a mixed population with pUC18, before and after a 12 day incubation in umDC- cells. p181.
Figure 6.16: mean normalised competition results. p182.

Figure 7.1: A polyacrylamide gel showing the effects of SOS induction and polymerase mutation on $(CCG/CGG)_{24}$ TR tract instability. p204.
Figure 7.2: A polyacrylamide gel showing the effects of SOS induction and polymerase mutation on $(CAG/CTG)_{43}$ TR tract instability. p205.
Figure 7.3: The percentage of expansions, deletions and no net change to CGG_{24} tracts recovered from various backgrounds. p207.
Figure 7.4: The percentage of expansions, deletions and no net change of CGG24 tracts recovered from various backgrounds. p208.
Figure 7.5: The percentage of expansions, deletions and no net change of CAG43 tracts recovered from various backgrounds. p209.
Figure 7.6: The percentage of expansions, deletions and no net change of CTG43 tracts recovered from various backgrounds. p210.
Figure 7.7: The location of Okazaki fragment termination and subsequent 5' flap processing could mediate TR tract instability. p227.
Figure 9.1: Mouse flanking sequences. p237.

List of tables

Table 1.1: The 9 expressed polyglutamine diseases recognised so far. p17.
Table 1.2: Human inherited disorders linked with non-coding TR tracts. p20.
Table 1.3: Genes for which up-regulated expression has been observed as part of the SOS response. p67.

Table 2-1: Strains of *E. coli* used in this work. p70.
Table 2-2: Plasmids used in this work. p74.
Table 2-3: Synthetic oligonucleotides for PCR and primer extension reactions. p75.
Table 2-4: Restriction endonucleases used in this work. p83.

Table 4.1: The mean arcsin values for each TR plasmid tested in each strain. p136.

Table 6.1: PvuII fragment size for each TR plasmid. p177.
Table 6.2: The mean values of plasmid loss or gain for each plasmid tested. p182.
Table 6.3: Sign test values for TR plasmid normalised competition results. p184.
Table 6.4: Outcomes of TR plasmid competition independent of TR orientation. p186.

Table 7.1: Proportions of deleted, expanded, and constant TR tract lengths in DNA repair mutants. p211.
Table 7.2: A comparison of average instability values for each TR plasmid. p212.

List of abbreviations, units and prefixes

Å, Angstrom

A, Adenine

C, Cytosine

CAG_{43}, pUC18 bearing a $(CAG)_{43}$ tract with mouse flanking sequences inserted into the multiple cloning site so that $(CAG)_{43}$ is the orientation encountered on the lagging strand during DNA synthesis. This same notation is used for all of the pUC18 plasmids bearing cloned TR tracts.

Ci, Curie

Conjugation, the union of two bacterial cells for the purpose of transferring chromosomal material from the donor to recipient.

°C, degree Celsius

Da, Dalton, a measurement of molecular mass, defined as $1/12^{th}$ the mass of a ^{12}C atom.

DNA, deoxyribonucleic acid

DNase, deoxyribonuclease

dNTP, deoxynucleotide triphosphate

DSB, double strand DNA breaks

DSBR, double strand break repair

dsDNA, double stranded DNA

DSE, double strand end. Usually applied to the duplex formed during replication fork regression by nascent strand annealing.

E. coli, Escherichia coli

g, gram

G, Guanine

Gene conversion, recombination in which one allele is used as a template to change a second allele into an homologous partner allele. The process may or may not involve crossover products.

HJ, Holliday junction

L, litre

Lagging strand, during DNA replication, the strand along which discontinuous synthesis progresses by ligating short (Okazaki) fragments synthesised individually in the 5'-3' direction, so as to follow the apparent 3'-5' direction of replication fork progression.

lawn, a continuous layer of bacteria on the surface of agar, used to grow bacteriophage.

Leading strand, highly processive DNA replication in the 5'-3' direction brought about by the continuous polymerisation at the growing 3' end.

m, meter

M, molar

mol, mole

MMR, mismatch repair

Pol, DNA polymerase

Prophage, a bacteriophage naturally inserted into a bacterial chromosome and capable of stable transmission down bacterial generations.

rec^{-}, displaying a recombination deficient phenotype.

RecA*, the nucleoprotein filament formed by polymerised RecA assembled on ssDNA, actively promotes strand exchange between homologous duplexes, and is responsible for stimulated LexA and UmuD autocleavage

Satellite DNA, DNA of repetitive sequence which forms a separate band in a density gradient after centrifugation.

SCD, the average fraction of supercoiled dimeric plasmid found in the total supercoiled monomer and supercoiled dimer plasmids extracted from a specific strain. (see page 131.)

ssDNA, single stranded DNA

T, Thymine

TR, trinucleotide repeat

Transduction, transfer of genetic material from a bacterial donor to a bacterial recipient using a bacteriophage as a vector.

Transformation, artificial modification of a genome by direct insertion into the cell of genetic material from another source.

TREDs, triplet repeat expansion diseases

u, unit

UV, ultraviolet light

WRN, Werner syndrome helicase

The Greek alphabet:

alpha,	β
beta,	α
gamma,	γ
delta,	δ
epsilon,	ε
zeta,	ζ
eta,	η
theta	θ
iota	ι
kappa,	κ
lambda,	λ
mu,	μ
nu,	ν
xi,	ξ
omicron,	o
pi,	π
rho,	ρ
sigma,	σ
tau,	τ
upsilon,	υ
phi,	φ
chi,	χ
psi,	ψ
omega.	ω

Prefixes:

Factor	Prefix	Symbol
10^{3}	kilo	k
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

Acknowledgements

Thanks must first go to everybody in the Leach group for their help, support, and shared interest in the topic of this thesis. Friends all. Thanks of course to David Leach; also to Gareth Cromie and John Connelly for brainstorming ideas and afterwork refreshments.

SPECIAL THANKS TO MY PARENTS FOR THEIR LOVE, AND CONSTANT SUPPORT AND ENCOURAGEMENT IN EVERYTHING I EVER DID. THE WORK PRESENTED IN THIS WEIGHTY TOME WOULD NOT HAVE REACHED THIS STAGE WITHOUT THE CONFIDENCE THEY INSTILLED IN ME.

This thesis is dedicated to my grandmas, Dorothy and Kathleen. My friends and benefactors.

Several members of the Edinburgh University Mountaineering Club have been great friends during my time here. Thanks for some awesome weekends on the ice and rock. Especially.... Big ups to Dan Cox and Naomi Towers. And those who didn't come back.

Stu and Deek deserve much more than a mention, but that's all they're getting because true friends don't need to be told. The same is true for Paul and Hannah.

Of course a special mention must go to Caroline, who, in the words of Bob Dylan, gave me shelter from the storm.

And finally, some advice to future students: " If ivir tha does owt fa nowt, do it for thisen"- nobody ever did a PhD for the money, so enjoy it.
However, remember....

"You can never conquer a mountain. If you are lucky, you will stand on the summit, but only for a few moments, then the wind will blow away your tracks."

Abstract

Expansion beyond a threshold length in specific trinucleotide repeat (TR) tracts have been shown to cause at least 15 inherited human neurological disorders. The mechanism of expansion of TR tract length is unknown, but is thought to relate to the unusual ability of (CAG/CTG) and (CCG/CGG) TR tracts to fold into stable secondary structures such as pseudo-hairpins. Hairpin structures within long palindromic sequences of DNA have previously been shown in *E. coli* to be sites of genetic instability mediated by hairpin cleavage and subsequent recombinational repair. TR tracts have also been shown to stall DNA replication forks *in vivo*, and recombination is known to be important in the re-initiation of stalled and broken replication forks. TRs have the ability to mis-anneal at any point along the length of the complementary strand tract, offering potential for changes in tract length during replication or homologous recombination. Therefore, the aim of the work presented in this thesis was to investigate the contribution of recombination to the instability of TR tracts carried in high copy number plasmids in *E. coli*.

Plasmid dimerisation was used as an assay for crossover production during recombination. Recent work has suggested biased Holliday junction resolution for different recombination substrates. In this way, the increased recovery of dimeric CTG_{43} plasmid compared to CAG_{43} and pUC18 control in *recF* cells was interpreted as the production of long ss overhangs in $(CTG)_{43}$ tracts, suggested to result from hairpin cleavage in ssDNA resulting from incomplete Okazaki fragment synthesis. The persistence of these and other recombination substrates was investigated using *sfiA-lac* and *λ-gal* assays. Both assays demonstrated that the SOS response was not de-repressed by the presence of $(CAG/CTG)_{43}$, or $(CCG/CGG)_{24}$ tracts in *E. coli*. However, differences in DNA maintenance between TR tract and standard plasmid DNA were detected after 12 day incubations following co-transformation with copies of pUC18 and pUC18 carrying TR tracts. Plasmids carrying short TR tracts of $(CCG/CGG)_{24}$, $(CAG/CTG)_{25}$, and $(CAG/CTG)_{28}$ displayed a surprising propensity to out-compete standard pUC18, interpreted as an ability of these tracts to initiate replication. However, plasmids carrying longer $(CAG/CTG)_{43}$ tracts displayed stark losses to standard pUC18, perhaps due to replication inhibition or elevated

frequencies of DSB formation. Finally, end-labelling of excised $(CCG/CGG)_{24}$ and $(CAG/CTG)_{43}$ tracts was used to examine tract instability in a variety of DNA replication, recombination, and repair mutants. All strains exhibited a greater propensity for deletion than expansion in all tracts. Small changes in tract length were far more common than larger changes. In general, the longer $(CAG/CTG)_{43}$ tracts appeared to be more unstable than the shorter $(CCG/CGG)_{24}$ tracts, and in both cases the orientation which displayed the greatest instability was the one in which the triplet sequence thought to form the most stable secondary structures was present on the lagging strand. TR tracts were still unstable in *recA*$^-$ cells, but $(CAG/CTG)_{43}$ tracts were more stable than in wild type cells. Other mutants found to destabilise TR tracts included *umuD*$^-$ (implicated in translesion synthesis and DNA checkpoint activities), the exonuclease activity of PolA (implicated in Okazaki fragment processing), and *holD*$^-$ (required for full processivity of replication forks). In addition, loss of SbcCD hairpin endonuclease specifically increased the instability of $(CTG)_{43}$ tracts present on the lagging strand, predominantly by elevating the frequency of tract expansions. Thus it would appear that TR tract instability can occur in any mechanism of DNA metabolism in which the tract is present as a single strand.

Chapter 1: Introduction

Triplet Repeat Expansion Diseases (TREDs)

An increasing number of human inherited disorders have been found to be associated with trinucleotide repeat (TR) tracts of sequence $(5'CXG3')_n$. The diseases share dominant inheritance and anticipation, that is too say the age of onset is reduced and the severity of the disease is increased in successive generations. Anticipation is caused by intergenerational increases in the length of the TR tract. The mechanism of expansion of the TR array at these disease alleles is unknown, and the possible involvement of recombination in mechanisms of TR tract instability is the subject of this thesis.

Approximately 10% of the human genome consists of DNA with highly repetitive sequences, but such sequences are generally not present in bacterial genomes, including *E. coli*. DNA microsatellites of short tandem repeats of 2-7bp occur 5-10 times more frequently in humans than equivalent random motifs (Han *et al.,* 1994; Ricke *et al.,* 1995). The biological significance of repeat sequences in the human genome may be structural (heterochromatin and nucleosome formation), regulatory (CpG islands) parasitic, or junk DNA. Sequence database searches using five or more CAG/CTG repeats have shown that a large number of human and eukaryotic transcription factors (including developmental TFs) contain CAG repeat tracts in their coding sequences (Bhandari and Brahmachari, 1995). Prokaryotic TFs on the other hand do not contain CAG repeats. It would seem that none of the suggested roles for repetitive DNA in eukaryotes are required in prokaryotes, so it may be that TRs are not tolerated in such an efficient genome as *E. coli*, where one round of replication already takes more time than the minimum required for a complete cell cycle in optimal growth conditions. TRs might have been selected against over many generations because of their suggested negative effect on genome replication, or any troublesome ability to compromise chromosome integrity, for example by the formation of double strand breaks (DSBs).

The 9 expressed polyglutamine diseases

Repeats causing the neurodegenerative conditions listed in Table 1.1 are all CAG repeats in coding sections of the genes which are translated and transcribed into polyglutamine stretches.

When expressed, these expanded TR tracts encode polypeptides which misfold and form antiparallel β-sheets of polyglutamine strands linked together by hydrogen bonds between their main chain and side chain amides, resulting in aggregates of ubiquitin-positive insoluble protein in cell nuclei (Perutz *et al.*, 1994). In polyglutamine disease patients the presence of intra-nuclear inclusions and subsequent cell death is observed in sub-populations of neurones in the central nervous system in a pattern specific for each disease.

Table 1.1: The 9 expressed polyglutamine diseases recognised so far.

Disease	Gene Product	Gene locus	Trinucleotide Repeat	normal TR length	expanded TR length leading to pathogenesis
Huntington's disease (HD)	Huntingtin	4p16.3	CAG	6-35 median = 19	40-121 mean = 45
SBMA[1]	Androgen receptor	Xq11.2	CAG	11-34	40-62
DRPLA[2]	Atrophin-1	12p13.31	CAG	3-36	49-88
SCA1[3]	Ataxin-1	6p22-23	CAG	19-36	43-81
SCA2	Ataxin-2	12q24.1	CAG	17-29	37-50
SCA3[4]	Ataxin-3	14q32.1	CAG	12-27	61-84
SCA6	CANA1A	19p13	CAG	4-18	21-30
SCA7	Ataxin-7	3p12-13	CAG	4-35	37-200

[1] SBMA is spinal and bulbar muscular atrophy (Kennedy's disease).
[2] DRPLA is detatorubral-pallidoluysian atrophy (Haw-River Syndrome).
[3] SCA are spinocerebellar ataxia types.
[4] SCA3 is also known as Machado-Joseph disease.

Intra-nuclear inclusions are hypothesised to resist degradation, prevent ubiquitin recycling, and/or disrupt the proteosome. Indeed, work by Cummings *et al.,* (1998) showed the colocalization of a proteosome and folding chaperone (HDJ-2/HSDJ) with ataxin-1 aggregates in affected neurones of SCA-1 patients and transgenic mice. Over-expression of HJD-2/HSDJ reduced ataxin-1 aggregates in HeLa cells, suggesting that this chaperone is able to recognise misfolded polyglutamine repeat proteins, enabling subsequent refolding or degradation. Other work has co-localised an array of chaperones, including Hsc70, Hsp70, NEDD8, and two Hsp40 family members (HDJ-1 and HDJ-2), with polyglutamine nuclear inclusions.

However, whilst neuronal intranuclear aggregates may be phenotypic of these diseases, one study in 1998 suggested that nuclear translocation (rather than formation of intranuclear aggregates *per se*) might be sufficient for development of neuropathogenesis. Klement *et al.,* (1998) observed nuclear import and disease symptoms but not intranuclear inclusion formation in mutant ataxin-1 SCA transgenic mice. However, Saudou *et al.* (1998) expressed recombinant Huntingtin containing a nuclear export signal, which inhibited the formation of intranuclear inclusions and reduced cultured neurone lethality.

Other work (Kim *et al.,* 1999) suggests that the formation of nuclear inclusions is part of a cellular defence strategy against mutant polyglutamine polypeptides. Caspases are cysteine proteases activated at the initiation of programmed cell death. Caspase-8 activation was found to be required for neuronal death in rats containing an expanded polyglutamine repeat (Sanchez *et al.,* 1999), and activated caspase-8 was isolated from an insoluble fraction extracted from the affected regions of brain in Huntington's disease patients. Several reports have implicated specific subsets of caspases in the cleavage of several polyglutamine neurodegenerative diseases, including Caspase-1, -2, -3, -7, -8, and -11, but it should be noted that only Caspase-8 has so far been shown to be specifically activated by expanded polyglutamine proteins. In addition, the rate of Huntingtin protein cleavage by apopain (a proapoptotic cysteine protease) is reduced for Huntingtin carrying expanded polyglutamine tracts (Goldberg *et al.,* 1996). The role of Caspases in neurodegeneration is currently under investigation, yet there remains a time discrepancy between the matter of hours required for

programmed cell death, compared to the development of human neurodegenerative diseases over many decades (see Evert *et al.*, 2000 for a review).

Intriguingly the expression in mice of a $(CAG)_{146}$ tract targeted into a random gene not correlated with any TR disease (hypoxanthine phosphoribosyltransferase HPRT) still induced nuclear inclusions and neuronal cell death (Ordoway *et al.*, 1997). It would appear that the presence of over 50 consecutive glutamine residues in any polypeptide is sufficient for neuronal toxicity. This may be connected to the observation that nucleosomes are preferentially positioned at long CTG repeat tracts (Wang, 1994).

Inherited disorders associated with unexpressed TR tracts

Expansions in the unexpressed trinucleotide repeat tracts linked with other neurodegenerative disorders (Table 1.2) seem to reduce expression of functional protein. In myotonic dystrophy, for example, a CTG repeat array interrupts a 3.5kb CpG island in the 3' untranslated region of the myotonic kinase gene. Expansion of this repeat tract beyond the normal 180-200 triplets eliminates an enhancer element that regulates transcription of the adjacent downstream gene termed DM locus-associated homeodomain protein (DMAHP) (Boucher *et al.*, 1995; Klessert *et al.*, 1997).

Table 1.2: Human inherited disorders linked with non-coding TR tracts.

Disease	Gene Product	Gene locus	Trinucleotide Repeat	normal TR length	Pre-mutation length	expanded TR length leading to pathogenesis
Fragile-X syndrome	FMR-1 (FMRP)	Xq27.3	CGG	6-53	59-230	230-2000
Fragile-XE syndrome	FMR-2	Xq28	GCC	6-35	31-61	200-900
Friedreich ataxia	Frataxin	9q13-21.1	GAA	6-29	34-40	200-900
Myotonic Dystrophy	Myotonic Dystrophy Protein Kinase (DMPK)	19q13	CTG	5-37	50-80	80-1000
SCA-8	?	13q21	CTG	15-27	110-200	>200
SCA12	PP2A-PR55β	5p31-11	CAG	7-25	?	49-75

Fragile-X syndrome (FXS) is the most common form of inherited mental retardation, affecting about one in 1500 males and one in 2500 females (Richards and Sutherland, 1992). Normal fragile-X mental retardation gene product (FMRP) is thought to function as an inhibitor of translation for a subset of mRNAs it shuttles between the nucleus and ribosomes in dendritic spine cells in the brain (Feng *et al.*, 1997). Rather like Freidrich's ataxia (FRDA), this syndrome requires a pre-mutation state of 5-50 repeats before a disease-causing mutation gives rise to extremely large expansions of 200-2000 triplets, which converts the locus into a folate-sensitive fragile site at Xq27.3 (Sutherland, 1988). Expanded CGG repeats in the 5' untranslated region of the *FMR1* gene are hypermethylated (Figure 1.1), with methylation spreading to an upstream CpG-rich promoter element. This promoter lacks a functional TATA box or initiator element, and normally has an intrinsic bend, which is enhanced by the binding of α-Pal/HRF-1 and USF1/USF2 heterodimeric transcription factors. Hypermethylation in FXS alleles prevents binding of transcription factor α-Pal/HRF-1 and reduces the affinity of USF1/USF2 binding. A more permanent inactivation of *FMR1* is also achieved by the methylated residues acting as targets for MECP2, which is known to recruit histone deacetylases and other proteins co-localised with transcriptionally silent heterochromatin.

Figure 1.1 CGG tract expansion in the *FMR1* gene reduces transcription

1.1a: Transcription factor binding to the promoter region in the normal allele of FMR1.

1.1b: Methylation in the FMR1 promoter of the FXS allele prevents NRF-1 and Sp1 binding.

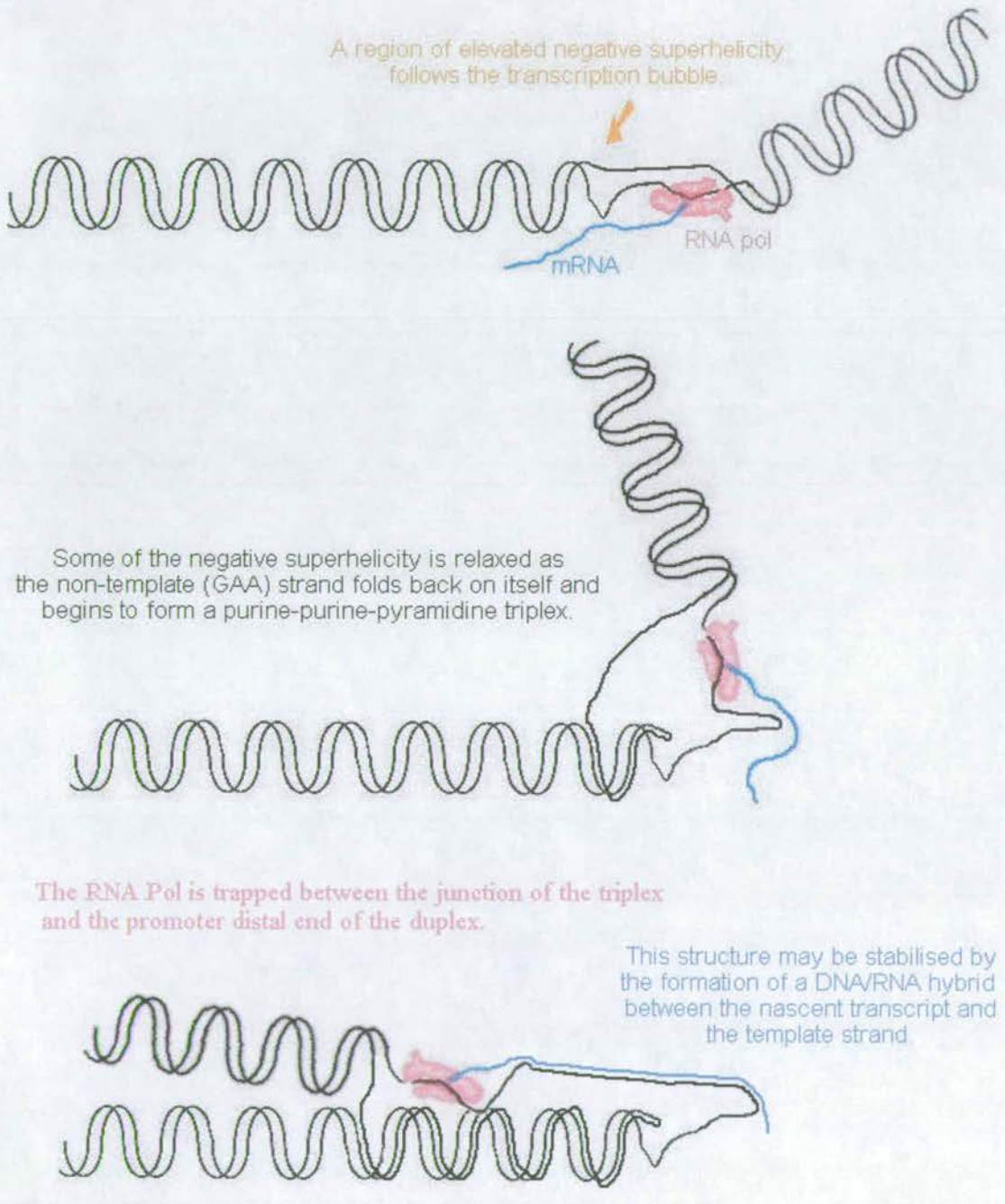
(adapted from Grabczyk *et al.*, 2001).

Freidrich's ataxia is unusual among trinucleotide repeat diseases in that it is autosomal recessive and is correlated with expansion of a GAA tract. The FRDA gene located on 9q13 encodes Frataxin, a mitochondrial protein of unknown function. The reduced frataxin levels found in FRDA patients is thought to cause mitochondrial iron accumulation and deficiencies in oxidative phosphorylation, particularly in sensory neurones and heart muscle. Expansion of a GAA repeat tract in the first *FRDA* intron is thought to reduce frataxin transcription elongation by trapping the advancing polymerase as the TR DNA forms a triplex structure (figure1.2). Expanded GAA repeats in nuclear extracts containing human RNA polymerase II experience difficulties in transcription elongation (Bidichandani *et al.,* 1998). Similar observations have been made with purified RNA polymerases, including that of bacteriophage T7. Most transcripts were truncated at the promoter distal end of the GAA repeats, suggesting that either the purine:purine:pyramidine triplex structure is extremely transient, or that a proximal duplex/triplex junction does not present a significant obstacle for the RNA polymerase. Interruptions totalling more than 20% of the GAA tract introduced enough mismatches to destabilise the triplex structure and prevent inhibition of *in vitro* transcription (Sakamoto *et al.,* 2001). In another cell-free system an oligonucleotide with a sequence designed to inhibit transcription-coupled triplex formation substantially increased the yield of longer transcripts (Grabczyk and Usdin, 2000).

It is significant that the biochemical properties of the unusual GAA repeat tract in Freidrich's ataxia favour folding of the tract into a triplex rather than hairpin or quadruplex formation (see below), and this has resulted in the unique disease mechanism of RNA polymerase trapping.

Figure 1.2 RNA Polymerase entrapment due to triplex structures within the FRDA gene

RNA polymerase becomes trapped between the duplex DNA and a triplex structure generated within the FRDA gene. A region of negative supercoiling that follows RNA polymerases is believed to help form the structures.



Adapted from Grabczyk and Usdin, 2000.

Folding properties of TRs

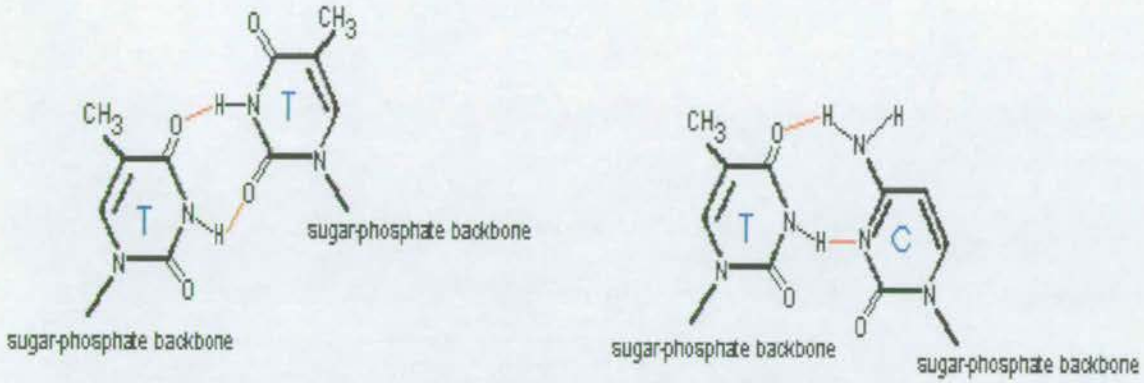
TR tracts present in single stands of DNA are able to form hairpins (intra-strand secondary structure) provided their stability is above a threshold of approximately-50kcal per hairpin (Gacy *et al.*, 1995). High resolution NMR has shown that single stranded CXG repeats as short as 12 residues have a high propensity for folding into anti-parallel duplexes (Zheng *et al.*, 1996). UV spectroscopic studies confirmed that CXG disease-causing repeats (and CTG in particular) exist in hairpin/duplex equilibrium due to the easy melting of complementary duplexes enabling the formation of homo-stranded structures. The central region of ss $(CGG)_{15}$ is susceptible to cleavage by $KMnO_4$/piperidine or P1 nuclease, suggesting a loop region of ss DNA at the apex of the hairpin (Mitas *et al.*, 1995). Gene fragments containing CTG or CGG repeats migrate faster than random DNA fragments of equivalent length in polyacrylamide gels under non-denaturing conditions (Chastain *et al.*, 1995; Chastain and Sinden, 1998). The sequence of the triplet repeats, the length of the repetitive DNA tract, and the flanking sequences of the locus determine hairpin stability. Thermodynamic and kinetic studies using UV spectroscopy have revealed that instability occurs at a much higher frequency at longer trinucleotide repeat tracts because of the extended lifetimes of long hairpins, but that hairpins were detected in repeats as short as $(CAG)_{25}$, $(CTG)_{25}$, and $(CGG)_{25}$ (Gacy and McMurray, 1998). Standard PCR through CGG repeats results in premature termination products, thought to occur as *in vitro* DNA synthesis stalls at intra-strand secondary structures formed when potassium is present in the reaction mixture (Usdin, 1995; Woodford, 1995).

Alternative structures in duplex DNA formed within the TR tracts of the myotonic dystrophy and fragile X loci are expected to contain mismatched base pairs (Pearson, 1997). Thymine has the ability to form non-Watson Crick hydrogen bonds with another thymine or cytosine, stabilising certain mismatched base pairs (figure 1.3).

The ability of TR tracts to form secondary structures is crucial to their mechanisms of expansion in tract length. In yeast, TR tracts capable of forming hairpins (CTG, CAG, CCG, and CGG) display 2 to 600 times the frequency of tract expansions compared to

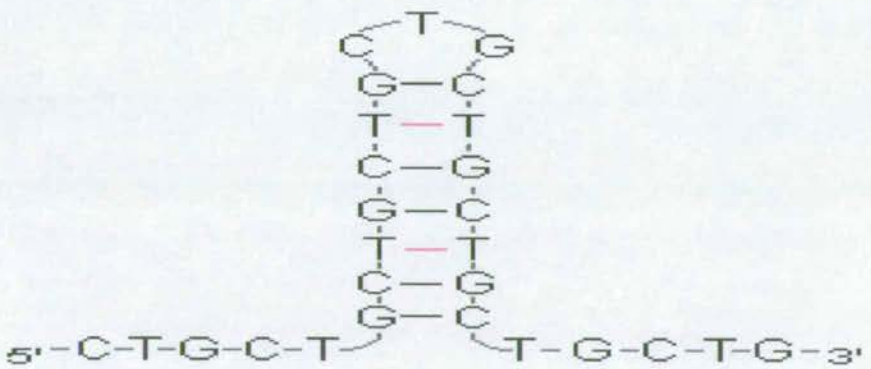
mononucleotide tracts or repetitive DNA that cannot form secondary structures (Spiro *et al.*, 1999).

Figure 1.3: Non Watson-Crick binding capabilities of thymine.



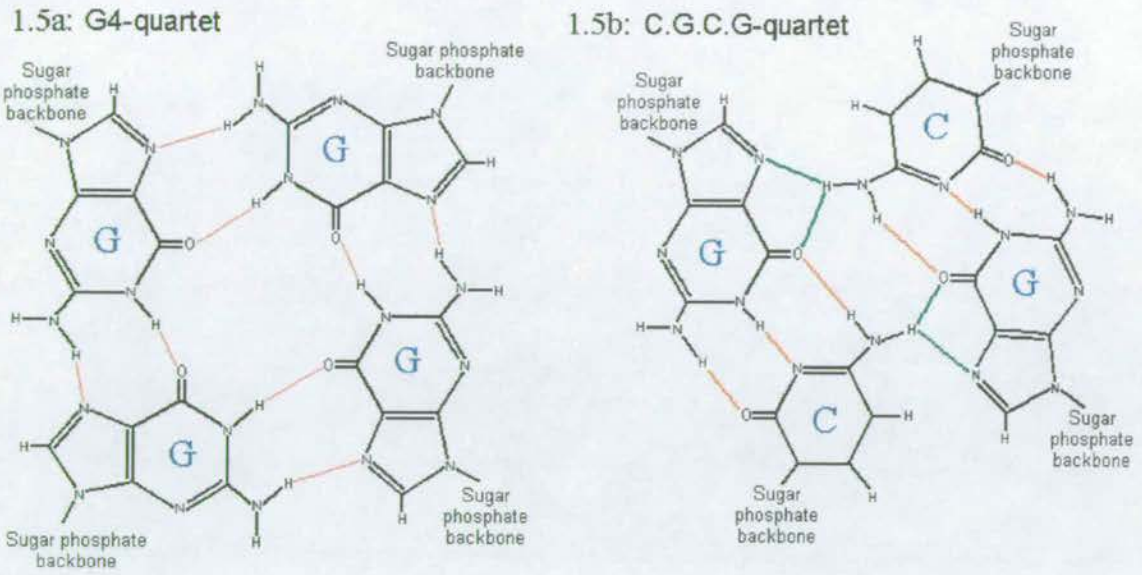
This non-Watson-Crick bonding occurs via hydrogen bonds, which increase the stability of intra-strand structure formation in CTG tracts compared to CAG tracts. This theorised difference in stability has supported the idea of hairpin formation (figure 1.4) in TR tracts *in vivo* by the observation of orientation-dependent expansions and large deletions in *E. coli* (Kang *et al.*,1995; Shimizu *et al.*,1996) and in *S. cerevisiae* (Maurer *et al.*,1996; Miret *et al.*,1998).

Figure 1.4: One form of hairpin structure predicted to form within a CTG repeat tract.



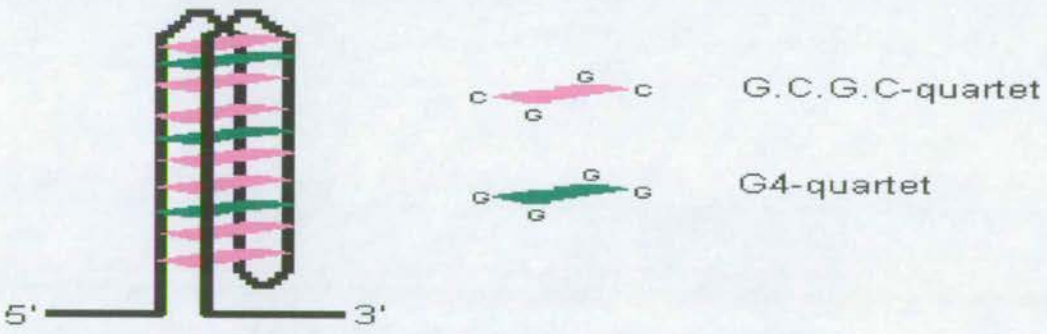
Similarly, guanine is able to form non-standard hydrogen bonds with another guanine (figure 1.5). This enables single stranded CGG tracts to form more stable structures than CCG tracts. One structure makes use of the ability of four guanines or two guanines and two cytosines to pair together in stable quartets.

Figure 1.5: Guanine has the ability to form four-base quartets.



Thus CGG repeat tracts are uniquely able to fold into quadruplex (tetrahelical) structures (Fry and Loeb, 1994; Usdin 1998), which can be thought of as folded hairpin structures (figure 1.6).

Figure 1.6: Quadruplex DNA composed of G4-quartets and G.C.G.C-quartets may form in CGG repeat tracts.



However the presence of a single loop in folded $(CGG)_n$ oligomers, and the accessibility of every guanine to dimethyl sulfate modification, suggests that hairpin rather than quadruplex structures are dominant (Nadel *et al.*, 1995).

In vitro studies have demonstrated stable hairpin formation in (CAG/CTG) tracts after duplex denaturing and renaturing (Pearson *et al.*, 1998), giving slipped-strand DNA (S-DNA) which migrates slowly through 4% polyacrylamide gel. This is possible in two complementary strands containing an equal number of repeats, as well as during

the annealing of two strands of TR tracts of different length (Pearson *et al.*, 2002). The propensity for S-DNA formation and the complexity of the structures produced increases with TR tract length (Pearson *et al.*, 1998). Electron microscopy has revealed that S-DNA structures (slippage loops) containing less than 10 triplets give bent DNA, whilst S-DNA containing between 15 and 50 triplets give multiple loops or hairpins, making complex structures such as three and four-way DNA junctions. The single-stranded character of these structures has been confirmed by their susceptibility to mung bean nuclease, and their branched nature by susceptibility to digestion by T7 endonuclease 1 (Pearson *et al.*, 2002). It is interesting that this study also revealed preferential binding of ssb (single strand binding protein) for (CAG) tracts rather than (CTG) tracts.

In vitro studies have demonstrated that formation of S-DNA structures can occur during DNA synthesis through (CAG/CTG) tracts. DNA synthesis using *E. coli* DNA polymerase 1 and human DNA polymerase β has been shown to pause after progressing 30-40 triplets into a (CAG/CTG) tract, and the stalled DNA synthesis products contained hairpin structures (Ohshima and Wells, 1997). This initial pausing of DNA synthesis by DNA secondary structure was thought to result in DNA slippage products as hairpins also form in the nascent DNA strand during spontaneous primer realignment.

In summary, the TR tracts implicated in genetic diseases are direct repeat sequences that have the ability to form folded DNA structures stabilised by non-Watson-Crick hydrogen bonding between bases. This characteristic is similar to the ability of inverted repeats (palindromes) to form hairpin and cruciform structures in DNA. Hairpins formed in palindromic DNA are substrates for SbcCD endonuclease, and are targets for recombinational repair. However, whilst palindromes do not appear to hinder replication, TR tracts have been reported to cause replication pausing.

The study of palindromic sequences has aided understanding of DNA hairpins and TR tract pseudohairpins

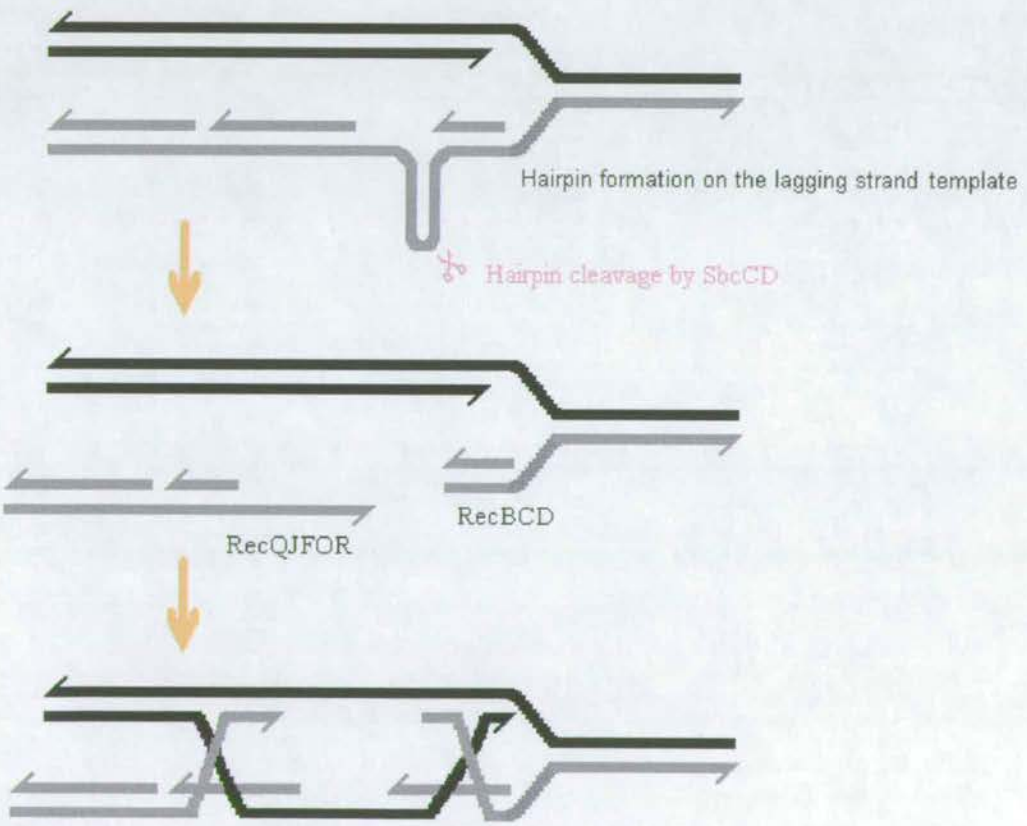
Palindromes are inverted DNA repeats that form hairpins or cruciform structures of definable length, with greater longevity than those formed in TR tracts. As such, their properties and requirements for secondary structure formation have been better documented than those of the more dynamic structures found in TR tracts.

Neither long TR tracts nor long, perfect palindromes (greater than about 200bp) are found in prokaryotes, but both are frequent within eukaryotic genomes (Wilson and Thomas, 1974). Short palindromes less than 40bp are common in regulatory regions of genes. They are thought to aid control of gene expression and types of DNA replication by providing distinctive DNA substrates for protein binding, and simultaneously causing sterically advantageous DNA bending and localised melting of the DNA duplex. DNA palindromes are also implicated in origins of DNA replication in the *E. coli* chromosome (Hirota *et al.*, 1979), prokaryotic plasmids (Lin, 1987; Wang, 1993), and near to the λ origin of replication (Hobom, 1979).

Hairpin formation is dependent on the DNA being single stranded, so could occur during many processes who's fidelity relies on the complementarity of DNA. This includes replication and certain forms of repair. In *E. coli,* loss of DNA bearing palindromes only occurs when the DNA is active (Leach and Lindsey, 1986), interpreted as replication-mediated increases in hairpin and cruciform extrusion due to transient elevations in negative superhelicity at single stranded DNA intermediates. Hairpin formation has not been demonstrated to occur during transcription, possibly because the transient length of ss DNA is too short. During replication of DNA, discontinuous synthesis on the lagging strand leads to long tracts of ss DNA between Okazaki fragments, and it is for this reason that hairpin formation is thought to occur more frequently on the lagging strand. Indeed, deletion of palindromic DNA mediated by replication strand slippage (see later) between direct repeats have been shown to occur on the lagging strand of replication in several systems in which the orientation of replication is known (Pinder *et al.*, 1998).

In *E. coli* the deletions observed in long artificial palindromes reduces their ability to form secondary structures, either by shortened length, or by introducing central asymmetries. Palindrome-mediated instability is not restricted to *E. coli*, as it has also been observed in *Bacillus* (Peeters *et al.*, 1988), *Streptococcus* (Behnke *et al.*, 1979), *Streptomyces* (Kieser and Melton, 1988), and *Saccharomyces cerevisiae* (Henderson and Petes, 1993; Gordenin *et al.*, 1993; Ruskin and Fink 1993; Nasar *et al.*, 2000). Mechanistic models for these deletion events can be divided into either replication slippage (figure 1.9), or cleavage of the secondary structure followed by recombinational repair (figure 1.7). Both of these models have also been proposed to mediate TR instability.

Figure 1.7: Proposed hairpin cleavage and recombinational repair based on genetic evidence from studies using long palindromes.



taken from Cromie and Leach, 2000.

Hairpin-mediated instability and inviability

In *E. coli*, long palindromes exhibit a *cis*-mediated inhibition on the net accumulation of replicated DNA molecules (Lindsey, and Leach, 1989). However, it is unclear whether replication itself is slowed across palindromes. It seems likely that replication forks are able to by-pass hairpins, but the hairpins remain and are susceptible to cleavage, with the subsequent risk of more extensive DNA degradation. A long palindrome (246bp) introduced into the *E. coli* chromosome by a phage λ lysogen has been demonstrated to be a hotspot for recombination (Cromie *et al.*, 2000). In keeping with this, recent work in yeast by Virginia Zirkian and co-workers has shown that TRs elevate recombination rates dependent on double strand break repair (DSBR) activities (Balakumaran et al., 2000). Palindromes in yeast also act as hotspots for meiotic recombination in the form of DSBR (Nag and Kurst, 1997; Nasar *et al.*, 2000), and the DSBs are thought to occur as a result of the action of a hairpin-specific endonuclease.

Similarly, palindrome-mediated instability and inviability (lack of successful transformants) in *E. coli* has been linked to recombinational DSBR (Leach *et al.*, 1997) as *E. coli* containing a cloned 246bp inverted repeat on their chromosome are inviable if mutated in *recA, recB,* or *recC* (see Recombination in *E. coli*, page 35). However, an additional mutation in *sbcC* suppresses this lethality, suggesting that SbcCD nuclease binds to the DNA hairpin formed by the palindrome, and causes a double strand break in the DNA, which must subsequently be repaired by recombination (Leach *et al.*, 1997). In *recA⁻sbcC⁻* cells containing the palindrome, a requirement for *recQ* was observed, suggesting an helicase-dependent pathway allowing replicative by-pass. PriA was not required, suggesting that although lagging strand synthesis may be halted (leaving a ss gap), progression of the replication fork itself is not hindered, and does not require fork re-initiation (Cromie *et al.*, 2000). This would suggest that generation of the DSBs occurs after replication fork progression through the palindromic DNA. In addition, the genetic requirement for both RecBCD (for the blunt ds end), and RecF (for the long ss overhang resulting from incomplete Okazaki fragment synthesis), suggests the presence of two DNA ends in the recombination mechanism, indicative of a genuine DSB rather than a collapsed replication fork.

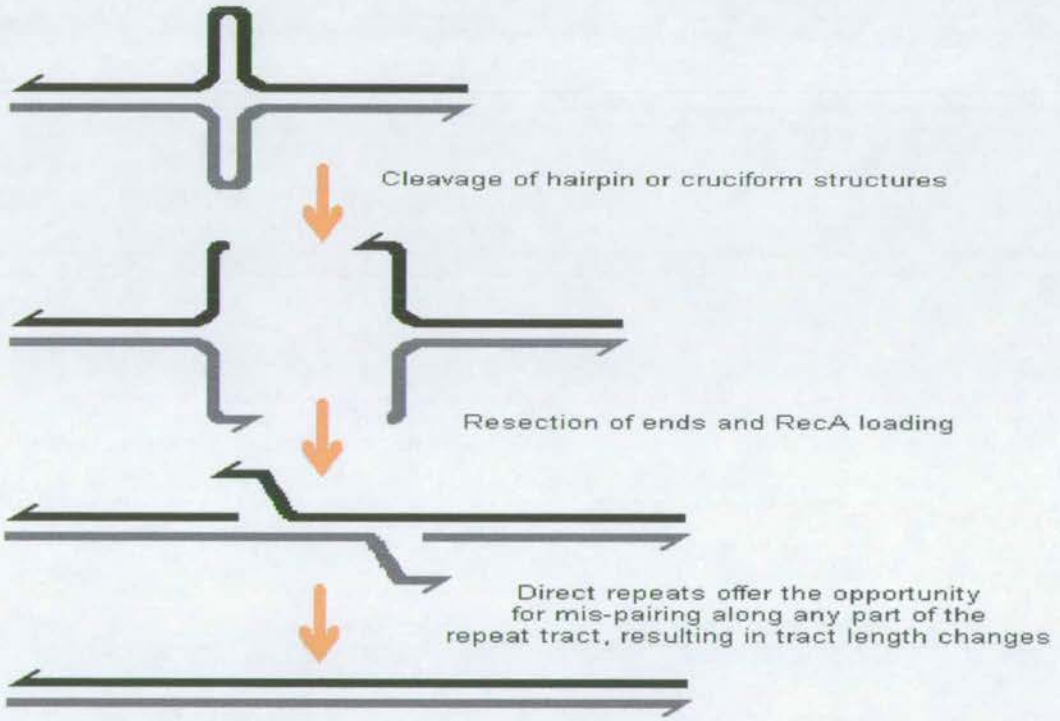
Together with the effects of hairpin processing outlined above, palindrome and TR instability is also affected by replication. Palindrome inviability in *E. coli* is dependent on replication of the palindrome (Leach and Lindsey, 1986; Shurvinton, 1987), which is consistent with both transient increases in negative supercoiling, and areas of single stranded DNA. Because replication on the lagging strand is discontinuous, single stranded DNA is likely to be more persistent and more common. Indeed elegant *in vivo* studies using asymmetric distribution of short direct repeats (one inside the palindrome, another outside the palindrome) described palindrome-related deletions to have occurred fifty times more frequently when replicated on the lagging strand (Trinh and Sinden, 1991; Rosche *et al.,* 1995; Sharples and Leach, 1996; Stukenberg *et al.,* 1994; Pinder *et al.,* 1998). 3-8bp direct repeats inserted within and without the palindrome greatly facilitated deletions, suggesting the nascent strand dissociates from the first repeat when it is stalled by the palindrome's secondary structure, and simply re-anneals with any local complementary sequences. This process is recA independent (Lovett, 1993), and indeed Lovett and Feschenko (1996) considered this process to occur via a misaligned replication intermediate. Thus palindrome instability and inviability can be mediated by hairpin-stimulated replication slippage.

cruciform extrusion

Experiments performed to investigate which conditions favour extrusion of palindromic sequences into hairpin structures have also proved valuable in developing models for secondary structure formation in TR tracts. Because inter-strand hydrogen bonds must be broken, cruciform extrusion from double stranded DNA would face a much larger energy barrier than the energetically favourable formation of a hairpin in single stranded DNA (Gellert, 1983; Sinden *et al.* 1983). When cruciform extrusion was demonstrated *in vitro* (Lilley, 1980; Panayotatos and Wells, 1981), a requirement was found to destabilise the DNA helix using either thermal energy or negative superhelicity (Lilley, 1980; Panayotatos, 1981). Indeed, the ability of palindromes to extrude increases with the negative superhelical density of the DNA (Mizuuchi and Gellert, 1982; Lilley *et al.,* 1985), as cruciform extrusion stabilises the release of torsional stress. Schaeffer *et al.* (1989) used equilibrium helix melting theory to demonstrate that superhelical stress present in circular DNA molecules have long-

range effects on extrusion of cruciforms, as increased tension reduces the thermal denaturation required for duplex melting. Cruciform extrusion from $(dA)_{81}(dT)_{81}$ tracts present in a pUC19 vector was used to demonstrate that plasmid superhelicity could be increased as a result of osmotic shock, anaerobiosis, and inhibition of protein synthesis (Dayn *et al.*, 1991). Cruciform extrusion from palindromic sequences occurs up to fifty times more frequently in *topA*$^-$ mutants (Sinden *et al.*, 1991), a result which confirms the significance of negative superhelicity (regulated by topoisomerases and gyrases) on palindrome extrusion.

Figure 1.8: Recombinational repair of processed cruciforms or offset hairpins could lead to changes in tract length.

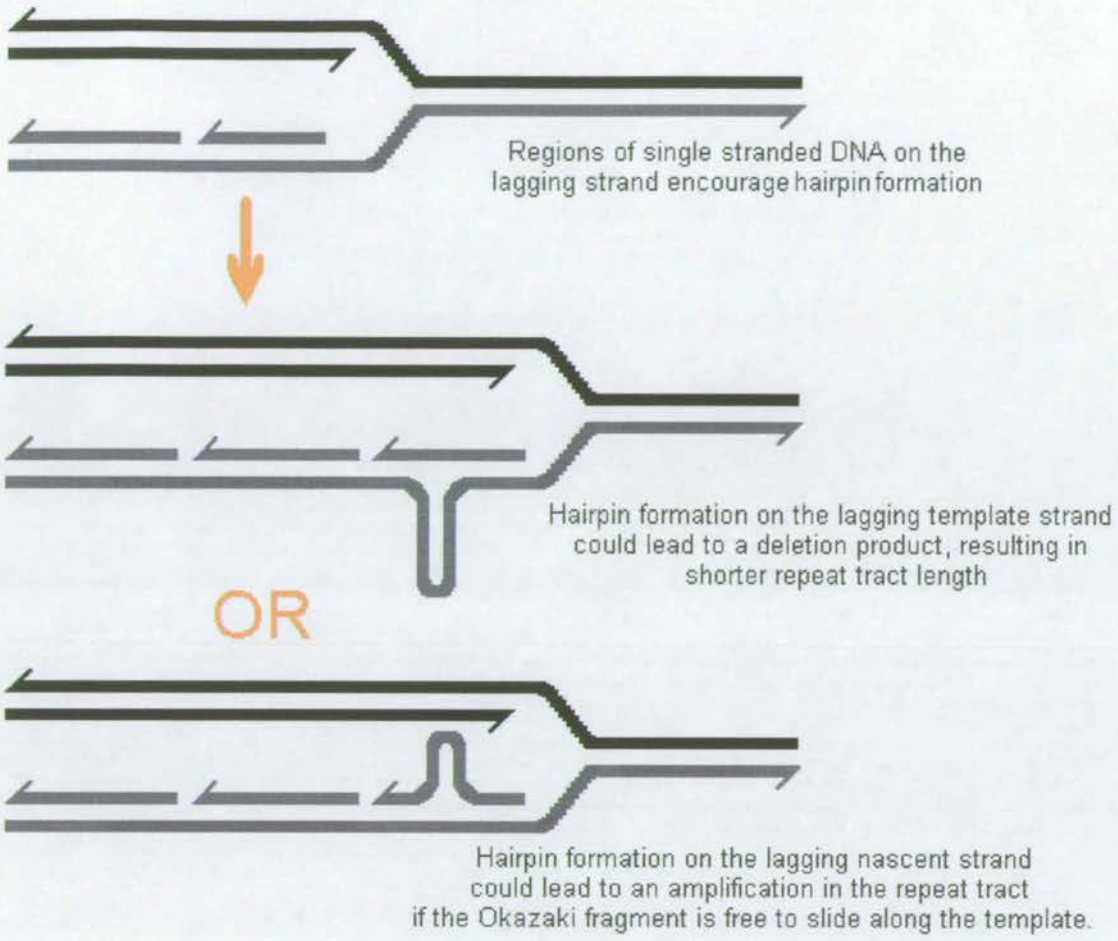


DNA melting pathways analysed by Ramstein and Lavery (1990), and base pair opening lifetimes measured in the region of 10ms at 21°C (Gueron *et al.* 1987), indicate that the double helical structure of DNA is indeed in dynamic equilibrium with other structures. Since DNA opens at the weakest site to reduce superhelical stress, TR tracts and palindromes have a higher propensity for secondary structure formation when located in an area of DNA which melts easily. When AT-rich sequences adjoin palindromes, the transition energy of cruciform extrusion is

significantly lowered, with longer AT-rich flanking sequences reducing the transition energy more than shorter AT-rich sequences (Wang and Sauerbier, 1989).

The preferred model of cruciform extrusion for the majority of inverted repeats involves an initial opening of basepairs limited to the centre of the inverted repeat, formation of intra-strand base pairing and a four-way junction, and finally branch migration to the fully extruded cruciform (Murchie and Lilley, 1987). This so called "S-type" cruciform extrusion requires cations *in vitro*, and predicts that mutations in the centre of the palindrome will be kinetically more influential on the extrusion rate than mutations further away from the centre of the palindrome. *In vivo* studies have confirmed that cruciform formation in *E. coli* can be altered by nucleotide sequence changes in the DNA lying between the inverted repeat sequences (Davison and Leach, 1994). This type of study was subsequently used to show that $(CAG)_2$, $(CTG)_2$, $(CCG)_2$, and $(CGG)_2$ TR sequences adopt stable and compact loops *in vivo* (Darlow and Leach, 1995). A 246bp perfect palindrome with a central GATC target for Dam methylase was cloned into bacteriophage λ and shown to be under-methylated *in vivo* and *in vitro* by its susceptibility to *Dpn*I cleavage (Allers and Leach, 1995). Dam methylase is unable to methylate its GATC target site if present as a single-strand, and a single stranded loop at the head of a stem is exactly what would be expected in a DNA hairpin.

Figure 1.9: Replication strand-slippage could lead to changes in tract length.



In contrast to the work outlined above which suggests that palindromes prevent the accumulation of DNA by endonuclease processing into DSBs, with the potential for subsequent degradation, TR tracts may be able to block passage of replication forks. G-rich repeat tracts of RNA and DNA that fold back to form hairpin or quadruplex structures have been shown to arrest DNA replication *in vitro* (Christiansen, Kofod and Nielsen, 1994). It seems likely that this is a consequence of a TR tract being able to fold at any point along its length and form stable structures. This is in contrast to palindromic DNA, which must have a long enough sequence to be stable, yet must also be single stranded along its entire length for the palindrome to be available for intra-strand pairing.

In longer tracts, multiple pairing possibilities allow rapid hairpin formation, and G-C rich hairpins are subsequently slow to dissociate (Gacy and McMurry, 1998). The extended lifetimes of a long hairpin may enable the polymerase to re-initiate replication, permanently trapping the structure. This would result in altered TR length

on one DNA strand unless a repair pathway is able to successfully resolve the hairpin. Thus strand slippage is a type of replication slippage stabilised by hairpin formation, and resulting in loops as the nascent strand is allowed to dissociate and re-anneal to any local complementary sequences.

In vitro replication through TR tracts has resulted in very large expansions, but expansions were only observed if the primer was complementary to the repeat tract rather than any flanking sequence (Lyons-Darden and Topal, 1999). This is interesting because it suggests a requirement during the replication slippage mechanism for a nascent DNA fragment to be free at both ends (figure 1.9), and hence able to slide freely along the repeat tract. This has been speculated to be an entire Okazaki fragment within the TR tract, or a prematurely terminated Okazaki fragment, or endonuclease action nicking a growing strand.

In summary, dynamic mutations might be mediated by strand slippage during replication, or recombinational repair of DSBs associated with secondary structure processing by endonucleases. By analogy with the secondary structures formed in palindromic DNA, it is possible to predict which factors might have an effect on secondary structure formation and dynamic mutation in TR tracts. We might expect long-range *cis*-effects of negative superhelicity, combined with short-range effects of AT-rich flanking sequences assisting duplex melting. Single stranded DNA (ssDNA) composed of short direct repeats has the ability to re-anneal to a complementary strand at any point in the repeat tract. Secondary structure formation should be far more likely in single stranded DNA, and a wave of negative superhelicity travelling behind a replication fork would greatly facilitate the formation of secondary structures in TR tracts (Gellibolian, Bacolla, and Wells, 1997).

Microsatellite Instability (MIN)

The problematic nature of loops formed in direct repeats, to DNA repair, has already been proven at microsatellite sequences by the disruption of the MutHLS mismatch repair pathway (MMR), responsible for correcting small loops of up to three nucleotides (Modrich, 1994, and 1997). Triplet loops can be generated by strand slippage during DNA replication (Streisinger, 1966). Microsatellites are composed of mono, di, tri, or tetra-nucleotide direct repeats. Since MMR always degrades the nascent strand (distinguished by a transient lack of methylation after replication in prokaryotes) and uses the original strand as a template for replication, it is a process which attempts to maintain the original length of the repeat tract. However, if short loops in a direct repeat tract are not recognised and processed correctly, they ultimately result in deletion (if a loop is formed in the template strand during replication), or amplification (if a loop is formed in the daughter strand during replication) of the repeat tract. In this way disruption of MMR results in small changes in microsatellite length called microsatellite instability (MIN). Mutation rates have been demonstrated to be inversely proportional to motif size (Charkraborty *et al.,* 1997), and the frequency of length changes in microsatellite tracts can be as high as 10^{-4}-10^{-2} events per cell per generation (Dallas, 1992).

$(CAG/CTG)_{64}$ tracts present in *mutS⁻ E. coli* have been shown to be very unstable, exhibiting a greater number of +1 and -1 triplet mutations in tract length (Schumacher *et al.,* 1998). Similarly, in yeast, mutations in *pms1* and *msh2* result in destabilisation of trinucleotide repeats by increasing the frequency of small amplification and deletion mutations, predominantly of only one repeat unit (Schweitzer and Livingston, 1997). Poly-(GT) repeat instability is increased 100 to 700 times in *S. cerevisiae* carrying mutations in MMR (Strand *et al.,* 1993). This shared phenotype of frequent, small changes in repeat tract length mirrors that of hereditary non-polyposis colorectal cancer (HNPCC) in humans. Approximately half of the HNPCC pedigrees studied so far are constitutionally heterozygous for hMLH1 (human MutL homologue 1) or hMSH2 (human MutS homologue 2), (Jiricny, 1998) and have functional MMR until the wild type copy is lost (consistent with the Knudson two-hit hypothesis, 1971).

It should be noted that loss of mismatch repair causes instability in the form of many small dynamic mutations at all types of microsatellite sequence across the genome. This is in contrast to the occasional, large amplifications found at specific TR disease alleles. However, it is possible that MMR may play some role in the processing of DNA structures formed in TR tracts, as they are likely to contain mismatched base pairs and small loops. Whilst MIN is clearly not the mechanism for single large expansions in TR tract length at specific disease alleles, it does demonstrate that the folding and strand slippage characteristics of TRs can lead to dynamic mutations if the pathway designated for their correction is dysfunctional.

Another DNA repair pathway, homologous recombination, is likely to encounter problems at trinucleotide repeat tracts. It is completely reliant on the complementarity of single stranded DNA intermediates, repairs a variety of DNA lesions, is intrinsically linked to DNA replication, and is responsible for large-scale genomic re-arrangement events when it goes wrong. Suppression of instability by the introduction of imperfect repeats into the TR tract is consistent with the observation that mismatches severely inhibit recombination (Chen and Jinks-Robertson, 1998).

Recombination in *E. coli*

In contrast to most other DNA repair systems that remove single stranded DNA lesions and use the remaining intact template for repair, recombination repairs lesions threatening the integrity of a duplex by actively finding an undamaged complementary strand to act as a template for replication across the lesion. Thus homologous recombination involves the exchange of identical or near-identical DNA between duplexes to prepare an intact template for accurate DNA replication through the damaged part of a genome. Recombination is used to repair several types of DNA lesion, all of which can be divided into two categories: double strand breaks (DSBs), and lesions present in single strand gaps (ss gaps). These substrates are generated naturally within cells as a result of endonuclease activities and replicative difficulties; they are generated by the external environment in the form of irradiation and chemical mutagens; and they are introduced into cells by conjugation, transduction, and artificial transformation with linear DNA. Recombinational repair of these substrates can generate genetic diversity by gene conversion, protein domain re-shuffling, horizontal gene transfer and meiotic gene shuffling. It is consequently a valuable catalyst of evolution.

Despite the variety of substrates for recombination, all reactions can be divided into a series of inferred, consecutive steps: presynapsis, synapsis, and postsynapsis (Clark, 1971). Presynapsis is an initiating step whereby a strand-exchange protein (RecA) is loaded onto the single stranded DNA of a recombination substrate (see figure 1.9). In synapsis, RecA mediates strand invasion into a homologous DNA duplex, and expels the homologous strand, forming a joint DNA molecule fused by a region of heteroduplex. Postsynapsis is an amalgamation of the final events in recombinational repair, including branch migration, resolution, and replication/ligation of any remaining gaps or nicks. Postsynaptic branch migration extends the heteroduplex region of the joint DNA molecule by catalysing strand exchange over a larger area, and gives rise to a four-way DNA junction referred to as a Holliday junction (after Holliday, 1964). Resolution of the four way junctions by cleavage of two opposite DNA strands enables the two DNA molecules to be separated. Replication and

ligation of any single stranded gaps or nicks within either molecule completes the process of repair.

Homologous recombination in wild type *E. coli* can be carried out by two main pathways, each involving a subset of proteins acting in a co-ordinated and sequential fashion on a specific DNA substrate. Recombination genes were originally divided into pathways by genetic analysis of recombinational repair proficiencies at specific DNA substrates in different genetic backgrounds, but these distinctions have become less rigid as research has progressed. Each recombination pathway was named after the enzymes used to carry out presynapsis (its initiating event). The RecB pathway is the dominant form of double strand break repair in *E. coli* (figure 1.10). The RecF pathway is thought to operate mainly in single-stranded gap repair (figure 1.11). A third, pathway, called the RecE pathway is only observed in *recBC⁻ sbcA⁻* cells carrying a cryptic prophage, and is only relevant in this thesis to the production of knockout mutants carried out in chapter 3.

The RecB pathway of double strand break repair

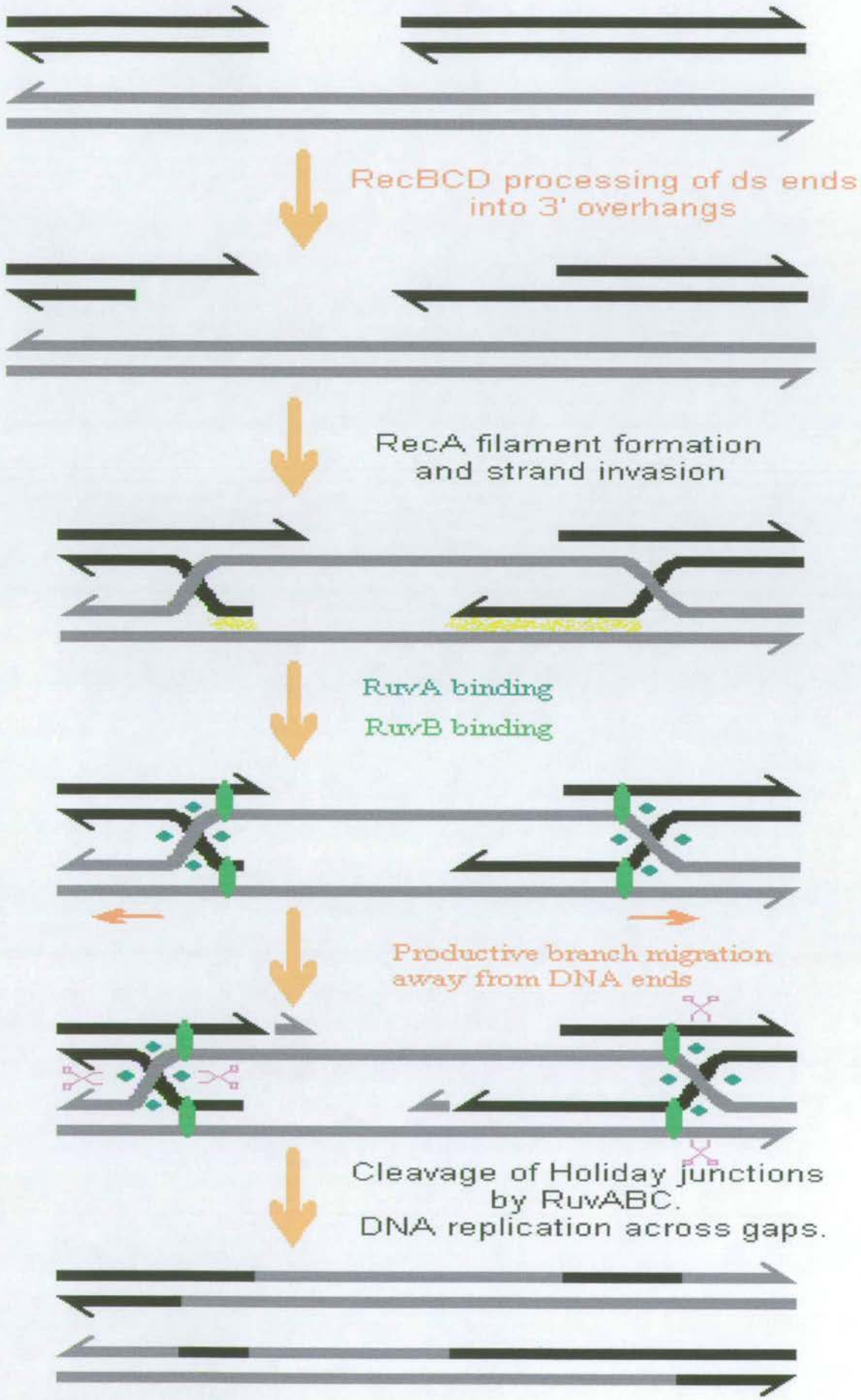
The RecB pathway was thought to be the dominant form of recombination in *E. coli* because *recBCD⁻* mutants have low residual levels of recombination as measured by generalised transduction, conjugation, and UV repair assays. However, these recombination events all involve DNA ends and subsequently the RecB pathway was seen as the dominant form of double strand break repair (DSBR) in *E. coli*. The importance of DSBR to *E. coli* is illustrated by reports of approximately 70% inviability of *recBCD⁻* cells in culture.

Accordingly, biochemical analysis has shown RecBCD (Exonuclease V) is a 330KDa heterotrimer which recognises blunt or nearly blunt ds DNA ends with less than 20 bases of overhanging ss DNA (Amundsen *et al.*, 1986). The recombinational deficiencies of *recBCD⁻* strains can be suppressed by cumulative mutations in *sbcCD*, and *sbcB* exonucleases, which allows the RecF pathway to initiate presynapsis from the persistent ss DNA overhangs.

Its ability to load onto blunt DNA ends (figure 1.10) coupled with its high processivity means that RecBCD is the major nuclease in *E. coli* for the degradation of linear double stranded DNA. Such destructive capability would be extremely dangerous for the genome of the host cell if it were not for the curious regulation of RecBCD's ds exonuclease activity. When RecBCD encounters a 5'-GCTGGTGG-3' sequence (referred to as a χ Chi site), a nick is created 4 to 6 bases 3' of χ, and the ds exonuclease activity is curtailed by inactivation of RecD, but the helicase and 5' ss exonuclease activity of RecBC carry the enzyme forward and actively load RecA onto the remaining 3'ss DNA end. Despite only 25-40% of RecBCD encounters with a χ site modifying the enzyme's behaviour (Taylor, 1992; Dixon and Kowalczykowski, 1993), most RecBCD-mediated recombination appears to be located around χ sites (Dower and Stahl, 1981), which are recombination hotspots within the *E. coli* genome. The genomic DNA of *E. coli* contains an enhanced concentration of these χ sequences (approximately once every 5Kbp), 90% of which are oriented toward *oriC*, protecting the *oriC*-proximal arm of the chromosome from degradation during DSBR of broken replication forks. *Chi* sites also effectively allow RecBCD to moderate its exonuclease activity based on differentiation between self and non-self DNA. So effective is this protection against viral DNA that bacteriophage λ has evolved an inhibitor of RecBCD called Gam, which is made use of in the engineered hyper-recombinogenic strain DL1218 used in this thesis (see chapter 3).

Whilst mutations in *recB* or *recC* reduce the frequency of conjugational and transductional recombinants by 100 to 1000 fold (Chaudhury and Smith, 1984), *recD* mutants are hyper-recombinogenic as a result of the ability of RecBC to unwind DNA ends and actively assist RecA loading onto the recombinogenic 3' end.

Figure 1.10: The RecBCD pathway of double strand break repair (DSBR).



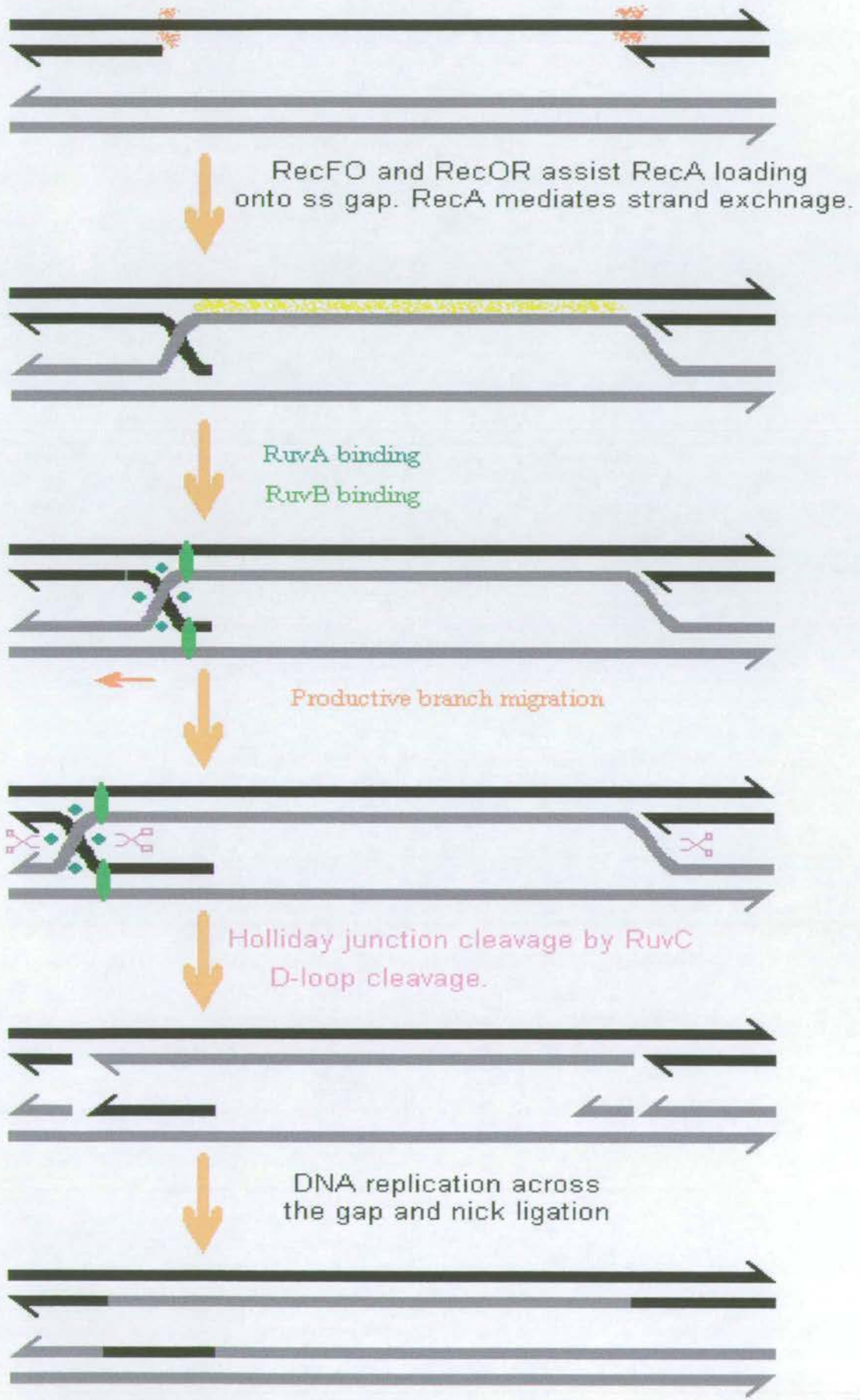
The RecF pathway of single strand gap repair

In *recBC*⁻ strains, residual recombination was found to be dependent on the presynaptic activities of *recF, recO, recR, recQ, recJ,* and *recN,* genes. RecF, RecO, and RecR form complexes which encourage RecA loading onto ss DNA, and limit its spread into ds DNA (figure 1.11). As such, the preferred substrate of the RecF pathway is single-stranded gaps, such as those generated during incomplete repair by other mechanisms (mismatch or nucleotide excision repair). Probably a more numerous substrate would be daughter strand gaps generated during DNA replication by one polymerase stalling at a DNA lesion, followed by re-initiation further along the template as the other polymerase carries the fork forward. Support for this comes from experiments showing that excision repair mutants are highly dependent on *recF, recO* and *recR,* for viability (Tseng, 1994), and that *recFOR* mutants exhibit UV sensitivity in *wt* cells. Mutations in *recF, recO,* and *recR* greatly reduce plasmidic recombination (James *et al.,* 1983; Laban and Cohen, 1981; Cohen and Laban, 1983; Kolodner *et al.,* 1985; Mahdi and Lloyd, 1989). Eukaryotes possess a direct equivalent to the RecF pathway in the RAD52 epistasis group.

The *recF*⁻ *recO*⁻ *recR*⁻ epistasis group

Many of the phenotypes of *recF*⁻, *recO*⁻, and *recR*⁻ are very similar. This epistasis group reflects the co-operative nature of the complexes formed; however, some distinguishing phenotypes will be discussed, along with biochemical signatures for each protein. Cells carrying mutations in *recF*⁻, *recO*⁻, or *recR*⁻ have reduced ability to repair daughter strand gaps (Rothman and Clark, 1977). Plasmid recombination is greatly reduced, but conjugational and transductional recombination is not significantly affected, except in a *recBC⁻sbcBC⁻* background (reviewed in Clark, 1991). SOS induction is delayed until one round of replication after exposure to harmful levels of UV light, but is then induced to a greater extent than *wt* cells (Hegde *et al.,* 1995; Whitby and Lloyd, 1995). This may be a consequence of replication forks breaking as they enter ss gaps and result in ds breaks. Over-expression of RecR in the presence of RecO can partially suppress *recF*⁻ related deficiencies (Sandler and Clark, 1994). In *wt* cells RecFOR are required for resumption of replication after UV damage (Courcelle *et al.,* 1997; Courcelle and Hanawalt, 1999).

Figure 1.11: The RecF pathway of single strand (ss) gap repair.



RecF

RecF is a 40Kda protein with affinity for the transition between single and double stranded DNA (Hegde, 1996). There are estimated to be less than 190 RecF monomers per cell. RecF transcription occurs as part of a DNA metabolism operon consisting of *dnaA, dnaN, recF,* and *gyrB*.

RecF is thought to target the anti-SSB activity of RecOR heterodimers to gaped DNA substrates (Hegde, 1996), aiding RecA loading. Indeed, non-null mutant RecA extracted from *recF*$^-$, *recO*$^-$, and *recR*$^-$ strains was found to be better at displacing SSB than RecA extracted from *wt* cells , allowing partial suppression of UV sensitivity (Volkert and Hartke, 1984). In the absence of RecOR, RecF has no effect on strand exchange at low concentrations, and actually inhibits exchange at high concentrations. Thus overproduction of RecF in a cell causes reduced viability and UV resistance. The RecFR heterodimer helps to prevent the spread of RecA filament from ss gaps into ds DNA (Webb, Cox, and Inman, 1997). *recF*$^-$ *priA*$^-$ cells (but not *recO*$^-$ *priA*$^-$ or *recR*$^-$ *priA*$^-$) are inviable (Sandler *et al.,* 1996), suggesting that RecF has an important function in replication fork re-initiation which partially overlaps that of PriA (see later).

RecR

RecR is a 22Kda protein capable of forming heterodimers with RecF or RecO. The *recR* operon contains *dnaX*, the gene product of which forms part of the γ complex of DNA Polymerase III, responsible for loading DnaN clamps onto primed DNA. Similarly, RecF is co-transcribed with DnaN, the beta subunit of DNA Polymerase III, which forms the 150 or so sliding clamps per cell for DNA replication. Because of this association with other proteins present at DNA gaps encoded by the *recF* and *recR* operons, it has been suggested that the RecFR complex might be deposited on the replisome proximal side of ss gaps by the replication machinery as it assembles DNA clamps at the start of Okazaki fragments (Webb, Cox and Inman 1997). This may aid RecFR to function in limiting the non-productive spread of RecA filament into ds DNA beyond daughter strand gaps.

RecO

RecO (RAD52) is a 26Kda protein with affinity for both ss and ds DNA. It is poorly expressed due to a weak promoter, a weak ribosome-binding site, and an abundance of rare codons. RecR monomers form heterodimers with RecO which promote recombination at ss gaps by favouring RecA binding of the SSB-coated ss DNA. RecOR does not displace SSB as such, but is able to bind SSB coated DNA using the RecO surface of the dimer.

The RecF pathway does not only repair ss gaps, but is also capable of repairing DNA ends. Indeed, RecOR dimers are thought to modulate RecA protein function at ss 5' DNA ends by helping to prevent filament disassembly (Bork, Cox, and Inman, 2001).

RecJ

RecJ (ExoIX) is a 60kDa 5' to 3' single-strand exonuclease (Kolodner, 1989). One presynaptic role for RecJ is the generation of 3' ss overhangs suitable for RecA loading from nearly blunt ended DNA. Since this activity seems to be mostly carried out by χ-activated RecBCD, only a very modest phenotype is observed in *recJ wt* cells (Lovett and Clark, 1984). However, conjugational recombination and UV resistance relies on RecJ in *recD*$^-$ strains (Lovett *et al.*, 1988; Lloyd *et al.*, 1988).

Another presynaptic role for RecJ is to extend the length of single stranded gaps in DNA, thus increasing the area available for RecA loading. This is achieved by the co-ordinated unwinding of DNA by RecQ helicase and simultaneous degradation of the 5' strand by RecJ. In *recB*$^-$ *uvrB*$^-$ cells, mutations in *recJ* cause severe deficiencies in ss gap repair (Wang and Smith, 1988). *recJ* mutants are extremely deficient in plasmid recombination (Kolodner *et al.*, 1985).

A postsynaptic role for RecJ has also been suggested in the catalysis of branch migration by imparting directionality on this process. As a recombinogenic 3' end invades another duplex during synapsis, the displaced complementary 5' strand could be degraded by RecJ, preventing competitive re-pairing with the other duplex strand (Freidman-Ohana and Cohen, 1998). RecJ also enhances RecA-mediated strand exchange and allows branch migration to traverse areas of non-homology *in vitro*

(Corrette-Bennett and Lovett, 1995). This is thought to occur by the coupled digestion of the expelled strand, which would simultaneously confer a 3' to 5' polarity on RecA-mediated strand exchange (Freidman-Ohana, Karunker and Cohen, 1999).

RecQ

RecQ is one of twelve helicases found in *E. coli*. RecQ is a 3' to 5' DNA helicase which can load onto blunt ends, but prefers 3' ssDNA overhangs (Umezu *et al.*, 1990). RecQ was first identified (Nakayama *et al.*, 1984) as a mutation in a *recBCsbcBC* background that caused a 100 fold decrease in recombination frequency, 20 fold increase in sensitivity to UV, and additional sensitivity to methyl-methane sulphonate and H_2O_2. Biochemical studies of RecQ demonstrated its function is analogous to RecBCD, but without exonuclease activity (Harmon and Kowalcykowski, 1998). SSB greatly stimulates the helicase activity of RecQ *in vitro* (Irino *et al.*, 1986).

recQ$^-$ strains exhibit similar phenotypes to *recJ* strains. Unlike *recF*$^-$, *recO*$^-$, or *recR*$^-$ strains, no defects in UV repair are observed in *recQ*$^-$ or *recJ* cells. Defects in conjugation and UV repair are only observed when combined with *recBC*$^-$ *sbcBC*$^-$ (Nakayama *et al.*, 1985), suggesting an involvement in the production of 3' overhangs from blunt ends. Indeed, *in vitro* work by Harmon and Kowalkzykowski in 1998 demonstrated that RecQ could initiate recombination from a blunt-ended duplex in the presence of RecA and SSB by providing 3' single stranded DNA for RecA loading. Interestingly, the same work showed that RecQ could disrupt illegitimate recombination by unwinding the joint molecules formed by short regions of micro-homology. Genetic evidence shows that a variety of *recQ*$^-$ mutants exhibit a 20-300 fold increase in rates of spontaneous illegitimate recombination and a 10 to 100 fold increase in UV-induced illegitimate recombination (Hanada *et al.*, 1997). Thus RecQ is able both to initiate homologous recombination, and suppress illegitimate recombination. Work in eukaryotic systems has provided genetic evidence for suppression of illegitimate recombination at replication forks by RecQ homologues (see chapter 3).

During the unwinding of DNA and strand invasion events in recombination, the regular negative DNA superhelicity of the *E. coli* chromosome must be maintained

within certain limits, otherwise the energy barriers for many aspects of DNA metabolism become too great. This crucial activity is performed by a combination of DNA gyrases and topoisomerases. One interesting property of RecQ is its ability to form a complex with Topoisomerase III and fully catenate ds DNA circles (Harmon *et al.*, 1999; Wu *et al.*, 1999). Harmon suggested that this combined activity of RecQ and TopoIII could regulate the levels of homologous and non-homologous recombination in *E. coli*, and others believe it to function more widely in the control of recombination, replication, or both (Chakraverty and Hickson, 1999).

Genetic evidence exists for the co-operative action of RecQ and RecJ in *E. coli*, as *recQ*⁻ mutants suppress the severe reduction in crossover frequencies of *recJ* mutants in a *recBC*⁻*sbcA*⁻ background, as well as suppression of UV, γ-ray, and H_2O_2 damage repair defects (Kusano *et al.*, 1994b). Eukaryotic homologues of RecQ also possess a 5' to 3' exonuclease domain equivalent to the activity of RecJ in *E. coli*. Members of the RecQ helicase family possess the ability to unwind certain DNA secondary structures implicated in trinucleotide repeat diseases. This aspect of RecQ action is discussed in chapter 3.

RecN

recN⁻ cells are constitutively active for SOS mutagenesis functions (Dunman *et al.*, 2000). This may result from an elevation in the amount of activated $RecA^*$ filament in the cell at trapped recombination intermediates (Simic *et al.*, 1991; Chua *et al.*, 1993). RecN is thought to function in the presynaptic processing of DSBs to produce 3' ends (Lloyd and Buckman, 1991). Certainly *recN*⁻ cells are sensitive to agents known to induce DSBs, including γ-irradiation, mitomycin C (Picksley *et al.*, 1984), and UV and X-rays (Sargentini and Smith, 1986). Deficiencies in conjugation were only observed when the *recN*⁻ allele was studied in a *recBC*⁻*sbcBC*⁻ background (Lloyd *et al.*, 1983), which led to its original classification as a member of the RecF presynaptic pathway. RecN is present at very low cellular concentrations, but its expression is greatly upregulated during the SOS response (Finch *et al.*, 1985). Little is known about the biochemical activities of RecN because it is hard to purify in a soluble form. However, RecN does have sequence similarities to eukaryotic Structural Maintenance of Chromosomes (SMC) proteins, which have molecular hinge domains separating

components of an ATPase active site, presumably operable when the hinge closes (Cobbe and Heck, 2000).

SbcCD nuclease

Mutations in *sbcC* or *sbcD* do not appear to have any effect on recombination proficiency in *wt* cells (Lloyd and Buckman, 1985). SbcC is a 118kDa protein, and also a member of the SMC family. SbcD is a 45kDa phosphoesterase believed to be a nuclease (Sharples and Leach, 1995), and is transcribed from the same operon as the SbcC protein (Naom *et al.*, 1989). The overlapping *sbcC* and *sbcD* genes were originally identified as suppressers of recombination deficiencies in *recBC*$^-$ cells. SbcCs associate with SbcD to form a 1.3MDa complex consisting of a coiled-coil rod joining two "head" domains (Connelly, 1998). Biochemical studies of this complex describe an ATP-dependent double-strand 3' to 5' exonuclease capable of translocating and digesting from either end of a linear DNA molecule (Connelly, 1999). It is also an ATP-independent single stranded endonuclease, and the absence of this activity is thought to divert recombination into the functional RecF pathway in a *recBC*$^-$ *sbcB*$^-$ *sbcCD*$^-$ background, as SbcCD normally degrades ss overhangs (suitable for RecFOR-mediated recombination) into blunt ends (preferred by RecBCD).

However, perhaps the most relevant property of the SbcCD complex to this study is its ability *in vitro* to bind to the closed ends of hairpin molecules and cleave the central loop at its 5' side (Connelly *et al.*, 1998 and 1999). This result supports the work described earlier concerning the involvement of SbcCD in processing mis-folded DNA *in vivo*. The human homologues of *sbcC* and *sbcD* are the *RAD50* and *MRE11* genes respectively, which are thought to transcribe elements of a similar complex implicated in eukaryotic DSBR.

RecA is the agent of synapsis in Homologous Recombination.

Because RecA plays a crucial role in homologous recombination, it was identified very early by its severe phenotype in conjugational recombination and UV repair (Clark and Marguilies, 1965). The recombination-deficient phenotype was extended to transduction (Hertman and Luria, 1967), plasmid recombination (Laban and Cohen, 1981; Hobom and Hogness, 1974; Fishel *et al.*, 1981), X ray and γ radiation repair (Howard-Flanders and Theriot, 1966; Krasin and Hutchison, 1977; Sargentini and Smith, 1986), and repair of ss gaps generated by excision repair (Youngs *et al.*, 1974; Smith and Sharma, 1987). Thus RecA has been implicated in recombinational repair of all types, regardless of whether the initial substrate was a ds break or a ss gap. Slow growth is observed in *recA*$^-$ cultures since up to 50% of the cells are inviable and up to 20% of the total DNA within these cultures is undergoing degradation at any one time. Of course, *recA*$^-$ strains are extremely sensitive to all types of DNA damage.

RecA is a 38KDa protein capable of binding both ss and ds DNA (Sancar *et al.*, 1980; McEntee *et al.*, 1981). The usual concentration of 1,000 to 10,000 RecA monomers per cell is increased by fifty-fold upon early SOS induction. It forms long polymers of RecA/DNA, producing a nucleoprotein filament right-handed helix composed of about 3 bases per RecA monomer (Stasiak and Di Capua, 1982; Di Capua *et al.*, 1982), and six RecA monomers per turn, which effectively extends the DNA to 1.5 times its normal length (Flory *et al.*, 1984). RecA is intrinsically capable of loading onto ss DNA in a 5' to 3' polar manner by co-operative binding (Griffith *et al.*, 1984; Register and Griffith, 1985). This occurs at a rate of 30 to 40 RecA monomers per second, which translates to approximately 100 ss nucleotides per second. Since RecA disassembly occurs in the same polar direction as assembly, a 3' to 5' treadmilling effect is hypothesised. Thus 3' ends are coated more readily by RecA filament, and hence are likely to be more invasive as recombination substrates than 5' ss overhangs (Konforti and Davis, 1990). RecA recycling occurs in an ATP-dependent manner, with hydrolysis to ADP lowering the affinity of RecA for DNA (Menetski *et al.*, 1990). SSB acts co-operatively by removing secondary structures in the ss DNA, aiding the formation of longer presynaptic filaments (Kowalczykowski, 1987). RecA

is capable of forming filaments on uniform double stranded DNA *in vitro* (Kowalczykowski and Krupp, 1987), but much more slowly than when it is able to nucleate on a region of ss DNA and extend onwards into a region of ds DNA (West *et al.*, 1980). The RecA filament contains 2 DNA-binding regions: DNA1 and DNA2 (reviewed in Kuzminov, 1999). Each of these sites is able to bind two strands of DNA, but a total of three strands per RecA filament is preferred. DNA1 is filled by the sugar phosphate backbone of the strand on which the filament has assembled, with bases facing inwards. The strand screened for homology transiently binds DNA2.

The method by which RecA-coated DNA searches for homologous sequences is believed to occur by transient extension and unwinding of possible target dsDNA, enabling homologous pairing mediated by DNA-RecA filament (Gonda and Radding 1986, Rould *et al.*, 1992). *In vitro,* the RecA filament can form joint DNA molecules from substrates with a minimum of 8 consecutively homologous nucleotides (McEntee *et al.*, 1979). The search for homology is ATP-independent, and occurs at an estimated rate of one hundred to one thousand base pairs per second. Once homology has been recognised, RecA mediates local denaturing of the target duplex and exchange of the invading strand for its homologue. This is done by the single stranded DNA in DNA1 forming hydrogen bonds with its complementary strand in the recipient duplex. The homologous strand in the recipient duplex is displaced to a secondary binding site in the periphery of the RecA filament for extraction by SSB. Extension of this hybrid duplex is ATP dependent and proceeds in a 5' to 3' fashion for several kilobases. Extension may proceed into four stranded regions (with the displaced strands simply forming another hybrid duplex), but a single strand gap of at least 6 to 15 nucleotides is required for the initial loading of RecA and subsequent strand exchange. The exact mechanism and its requirement for ATP is not as yet fully elucidated. Menetski and Kowalczykowski (1990, and Menetski, Bear and Kowalczykowski, 1990) have suggested that ATP may be required for partial filament disassembly to aid reorganisation at points of heterology or where the filament enters a duplex region. Cox (1987), on the other hand, has suggested that ATP is used to drive strand exchange by rotating the DNA helices around their long axes.

RecA polymerisation is subject to several levels of control from multiple pathways of DNA repair. RecA polymerisation is promoted by RecFOR and RecBCD.

Transcription of *recA* is elevated as an early SOS response, ensuring positive feedback under cellular conditions where normal levels of recombination are inadequate. RecA filament formation is inhibited by MutSL and UvrD, allowing mismatch repair and UV excision repair to attempt repair if the substrate is suitable. The presence of MutSL inhibits RecA-catalyzed strand exchange between diverged sequences by preventing branch migration (Worth *et al.,* 1994). UmuD is also an inhibtor of homolgous recombination (Rehrauer *et al.,* 1998). RecA filament can also be dissociated by both RuvAB and RecG, suggesting the possibility of an internally regulated abortion mechanism within recombination pathways.

Postsynapsis

Postsynapsis consists of branch migration and cleavage of the Holliday junctions (HJs). Branch migration is used to extend the region of heteroduplex. Resolution of the HJs creates two recombinant duplexes that are physically independent of each other.

Branch migration is a thermally neutral process brought about by two base pairs simply swapping binding partners between two branches (Thompson *et al.,* 1976). However, branch migration of naked DNA under cellular conditions is thought to be quite slow because of the stacked X-structure of the HJs adopted in the presence of magnesium ions (Clegg, Murchie, and Lilley, 1994). RuvA binding forces HJs to adopt a more open conformation better suited to quick branch migration, and RuvB binding imparts directionality to migration. This allows branch migration to occur at several thousand base pairs per second. However, base mismatches impede branch migration by disrupting the helical stacked structure of 4-way junctions (Panyutin and Hsieh, 1993).

RuvABC

The RuvABC holoezyme is the major contributor to branch migration and HJ resolution in *E. coli.* Essentially, RuvA functions to open the HJ, RuvB operates as a translocase for branch migration, and RuvC is a nuclease which causes symmetrical breaks within the four-way junction.

The *ruv* genes were originally identified by their mitomycin C sensitivity (Otsuji *et al.,* 1974). *ruv*$^-$ strains are deficient in UV repair and plasmid recombination (Lloyd *et al.,* 1984). Defects in chromosome partitioning result in cells with long, nonseptate, multinucleate filaments (Ishioka *et al.,* 1998). This phenotype is relieved in *recA*$^-$ *ruv*$^-$ strains, pointing towards the postsynaptic action of the RuvABC complex. *ruvA* and *ruvB* are transcribed from the same operon capable of an eight-fold increase in the concentration of both proteins during SOS-induction. RuvC is transcribed separately and is not SOS-inducible. All three *ruv* genes have a very similar phenotype, suggesting that they functionally interact in an active complex. This hypothesis was supported by *in vitro* work showing that RuvA, RuvB, and RuvC form a complex together on HJs (Davies and West, 1998). Also, antibodies raised against any individual Ruv protein inhibited both branch migration and HJ resolution activities of RuvABC *in vitro* (Eggleston *et al.,* 1997).

RuvA is a 22Kda protein which forms a homotetramer. RuvA has a high binding affinity for four way DNA junctions (Parsons *et al.,* 1992), altering their structure into an open planar conformation (Parsons *et al.,* 1995; Rafferty *et al.,* 1996). One RuvA tetramer binds above a Holliday junction, and one binds below. The basal level of expression of RuvA is 700 molecules per cell.

RuvB is a 37kDa protein which displays low affinity for DNA, and is thought to be recruited to four way junctions by RuvA. RuvB forms a hexameric "doughnut", and in the presence of RuvA operates as a 5' to 3'helicase, as measured by displacement of single strands bound to ss circular DNA (Tsaneva *et al.,* 1993). RuvB is able to branch migrate on its own, but the optimal complex for this activity is made up of one tetramer of RuvA holding a Holliday junction "open", and two hexamers of RuvB on diametrically opposed duplex arms entering the junction (Muller, 1993). DNA moves through RuvA before entering the RuvB helicase rings (West, 1996). The basal level of expression of RuvB is 200 molecules per cell.

RuvC is a 19KDa endonuclease which forms dimers with binding affinity for Holliday junctions (Iwasaki *et al.,* 1991; Dunderdale *et al.,* 1991). *In vitro* studies by Bennett and West in 1995 showed that RuvC dimer binding to HJs opens the four way junction into a symmetrical structure with the DNA strands making up 2 wide and 2

narrow angles. RuvC cleaves Holliday junctions in a symmetrical manner by cleaving strands of the same polarity that make up the wide angles in the HJ. RuvAB has been shown to stimulate HJ resolution by RuvC (Zerbib *et al.,* 1998). A preferred sequence for cleavage at HJs is 5'-A/TTTG/C-3', with the cleavage occurring 3' to the run of Ts when this base is positioned at, or within one base of the crossover point (Dunderdale *et al.,* 1991; Iwasaki *et al.,* 1991; Bennett, 1993; Bennett and West, 1996).

In the presence of RuvAB, the cleavage-targeted strands have been identified as the ones that pass 5' to 3' through RuvB into the HJ (van Gool *et al.,* 1999). In this way, the generation of a crossover or non-crossover product of recombination is thought to be determined by the initial loading of RuvAB onto the synaptic substrate. In a crossover, the two recombining duplexes are physically spliced together to create novel arrangements of flanking genetic markers on the same strand, whereas non-crossovers retain the unmodified (parental) configuration. Different recombination substrates demand specific loading of RuvAB for productive branch migration, so various recombination substrates can be predicted to have preferred crossover or non-crossover outcomes. For standard recombination events, ends-in DSBR has been shown to result in crossover products, but ss gap repair would produce non-crossover products. Recombination initiated at replication forks by a leading strand break or paired nascent strands after fork regression, would result in non-crossover products, but recombinational repair of a break in the lagging strand would result in crossovers (Cromie and Leach, 2000; Seigneur *et al.,* 1998).

Crossovers are not as desirable an outcome as non-crossover products, as they can result in gene conversion and dimerisation events. Since RuvC cleavage of Holliday junctions is directed, it has presumably evolved to give non-crossover products in the majority of recombinational events encountered in the cell (Barre *et al.,* 2001). This is supported by the prediction that leading strand breaks in replication forks will not induce crossovers, whereas lagging strand breaks will produce crossovers. Leading strand replication would be most likely to encounter template gaps and generate DSBS because the leading strand template was the nascent lagging strand in the previous round of replication (Cromie and Leach, 2000).

An independent reaction used to reverse replication forks is carried out by a complex consisting of just RuvA and RuvB occurs at stalled replication forks (Seigneur *et al.*, 1998, see "recombination is needed for replication", page 56).

RecG

recG⁻ strains display moderate sensitivity to ionising radiation, UV light, and mitomycin C. In contrast to a 3 to 10 fold deficiency in conjugational recombination, plasmid recombination is unaffected. The residual level of recombination in *ruv* mutants is dependent on the product of the *recG* gene (Lloyd, 1991). *recG*⁻ and *ruv*⁻ single mutants are only moderately defective in recombination, but *recG*⁻ *ruv*⁻ double mutant strains are extremely deficient in recombination (Lloyd and Buckman, 1991). This suggests functional overlap in terms of branch migration and resolution between RecG and RuvABC. However, RecG does not possess endonuclease activity, nor does it interact with RuvC (Lloyd, 1991), so the mechanism of RecG-mediated HJ resolution is speculated to involve branch migration to a nick located elsewhere within the heteroduplex of the D-loop (Kuzminov, 1996). RecG is expressed at a low level which is not upregulated during the SOS response. *recG*⁻ strains display a 2 to 3 times greater expression of the SOS response genes than the normal basal level.

RecG is a 76Kda protein possessing a weak 3' to 5' DNA helicase activity (Whitby, 1994). RecG binds specifically to branched DNA molecules and unwinds them in an ATP-dependent manner (Whitby *et al.*, 1993; 1994; 1998). The branch migratory activity of RecG is unlike that of RuvAB in that it occurs in the opposite direction, is blocked by 30 nucleotides of heterology (Whitby and Lloyd, 1998; and Lloyd 1993), and prefers to bind and branch migrate three-way junctions rather than four-way junctions (McGlynn and Lloyd, 1999). This has led to the suggestion that RecG secures a D-loop strand invasion intermediate by migrating it into a duplex-duplex region, allowing subsequent RuvAB-mediated branch migration of the new four-way junction to a preferred RuvC cleavage sequence (Ohana, Karunker and Cohen, 1999). Whitby and Lloyd (1995) have suggested that the preference for three-way junctions may enable RecG to convert three-stranded substrates (such as stalled replication forks) into four-way substrates suitable for RuvABC resolution. This activity could function in stalled replication fork regression away from DNA damage or blocks caused by transcription (McGlynn and Lloyd, 2000). RecG is uniquely able to branch

migrate through DNA covered in RecA filament (Whitby, 1993). Since RecG is able to reverse RecA-driven strand exchange, it may bias the use of 3' DNA ends in recombination by reverse branch migrating from recombination initiated from 5' ends (Whitby and Lloyd, 1995).

Mutations decreasing the helicase activity of PriA suppress the repair and recombination defects of *recG* mutants, but not those defects in a *ruv recG* strain brought about by additional defects in HJ resolution (Al-Deib, Mahdi and Lloyd, 1996). Since PriA and RecG bind to similar but distinct forms of branched DNA (McGlynn *et al.,* 1997), it has been suggested that they compete for binding at recombination intermediates, and subsequently determine whether the intermediate is to be used to initiate replication (PriA), or to continue recombination (RecG) via the formation of Holliday Junctions (Al-Deib, Mahdi and Lloyd, 1996).

The loosening distinction between recombination and replication

Cromie and Leach (2000) suggested that Holliday junction resolution was biased in favour of non-crossover products in the majority of recombination events occurring at replication forks. In fact, recombination may occur at disrupted replication forks (estimated at one fork per cell per generation) more often than any other substrate in bacteria (Cox *et al.*, 2000). The requirement for DnaC in loading DnaB replicative helicase has enabled an estimation of the frequency of replisome re-assembly at collapsed forks in *E. coli*. An average of 18% of $DnaC^{ts}$ cells were unable to complete chromosomal replication (Maisnier-Patin, Nordström and Dasgupta, 2001). This suggests that only a small minority of replication arrest events (probably fork collapse events) require complete replisome re-assembly for resumption of replication, the rest being re-initiated without replisome dissociation. In mammals, ten forks undergo recombinational repair in every mitotic division (Haber, 1999).

The high frequency of replication fork arrest combined with observations of the requirement for recombination activities for their re-initiation has led to speculation that repair of stalled or broken replication forks is the primary function and evolutionary origin of homologous recombination systems in bacteria (Cox, 2001). There are several methods of repair of stalled forks, each involves a subset of recombination proteins acting at a series of distinct substrates at the disrupted fork. These substrates will be specific to which DNA lesions on which strand have disrupted the fork, and whether the entire fork has stalled, or whether the two polymerases have become uncoupled, with one advancing farther than the other. Since replication of DNA relies on complementarity, a DNA lesion present in one template strand may result in structures in which both DNA strands are affected. Repair of these structures is an obvious target for recombinational repair with its ability to utilise an undamaged complementary strand. The two substrates (breaks and gaps) of the repair pathways of homologous recombination can both be generated during DNA replication (DSBs by a fork running into a nick or gap in one template strand, and gaps by the stalling and re-initiation of a single polymerase carried forward by the replisome).

D-loop stabilisation converts a common recombination intermediate into a substrate for initiation of DNA replication

During homologous recombination, a transient D-loop is produced from a ss end in the invading strand (McEntee *et al.,* 1979). The structure of a D-loop is effectively an area of opened duplex with an invading 3'-OH strand of complementary DNA bound within this bubble. This structure strongly resembles the leading strand of a replication fork, with the paired 3' end acting as a primer. Recently, a priming pathway involving PriA has been shown to initiate long-range, stable DNA replication from these structures (McGlynn *et al.,* 1997). A D-loop capable of acting as a substrate for replication initiation must be a joint molecule formed by a stable interwound plectonemic joint (reviewed in Kowalczykowski, 1994). Recombination from single stranded gaps or between two duplexes produces unstable, non-interwound paranemic joints (DasGupta *et al.,* 1980). According to *in vitro* work by Cunningham *et al.,* (1980), a topoisomerase is required to convert these transient, unstable recombination intermediates into stable, catenated plectonemic joints. The stabilisation of D-loops allows RecA filament dissociation whilst maintaining the D-loop structure. This is necessary, because RecA filament would block polymerase access to the bound 3' end (Cox, 2001).

Replication initiation by PriA

PriA is intimately involved in re-priming broken replication forks. PriA is an 82Kda 3' to 5' helicase (Lee and Marians, 1989). The helicase activity of PriA is not required for the establishment of replication forks (Zavitz and Marians, 1992). A primosome is a multi-component complex required to prime DNA replication. It synthesises a short RNA primer bound to one strand of an opened DNA duplex, which acts as a substrate for assembly of the replication machinery (replisome). PriA was originally discovered as one of five proteins required to initiate assembly of a primosome onto the Primosome Assembly Site of θX174 for replication mediated by Polymerase III (review in Marians, 1992), but is now known as the "restart primosome". A primosome is required to move along the lagging strand of DNA replication (using DnaB replicative helicase) and synthesise RNA primers (using DnaG primase).

Primosome assembly at *oriC* is initiated by DnaA, and does not require PriA, PriB, or DnaT (Kaguni and Kornberg, 1984).

Mutations in the DnaA/*oriC* pathway produce more severe phenotypes than mutations in the PriA pathway of primosome assembly. However, *priA*$^-$ mutants defective in primosome assembly have massively reduced viability and rapidly acquire suppresser mutations in *dnaC*, another member of the primosome assembly complex (Lee, 1991; Sandler, 1996). P1 transduction is reduced 20 to 50 fold in *priA*$^-$ and there is a significant decrease in the frequency of generation of Hfr conjugation recombinants (Kogoma *et al.*, 1996). These "ends-out" recombination reactions are viewed as a form of replicative homologous recombination. PriA is thought to have a substantial role in DSBR regardless of which pre-synaptic pathways generate D-loops (Liu and Marians, 1999). This is clear from defects when *priA*$^-$ allelles are introduced into *recBC*$^-$ *sbcA*$^-$ and *recBC*$^-$ *sbcB*$^-$ strains, in which the RecE and RecF pathways respectively initiate DSBR (reviewed in Marians, 1999).

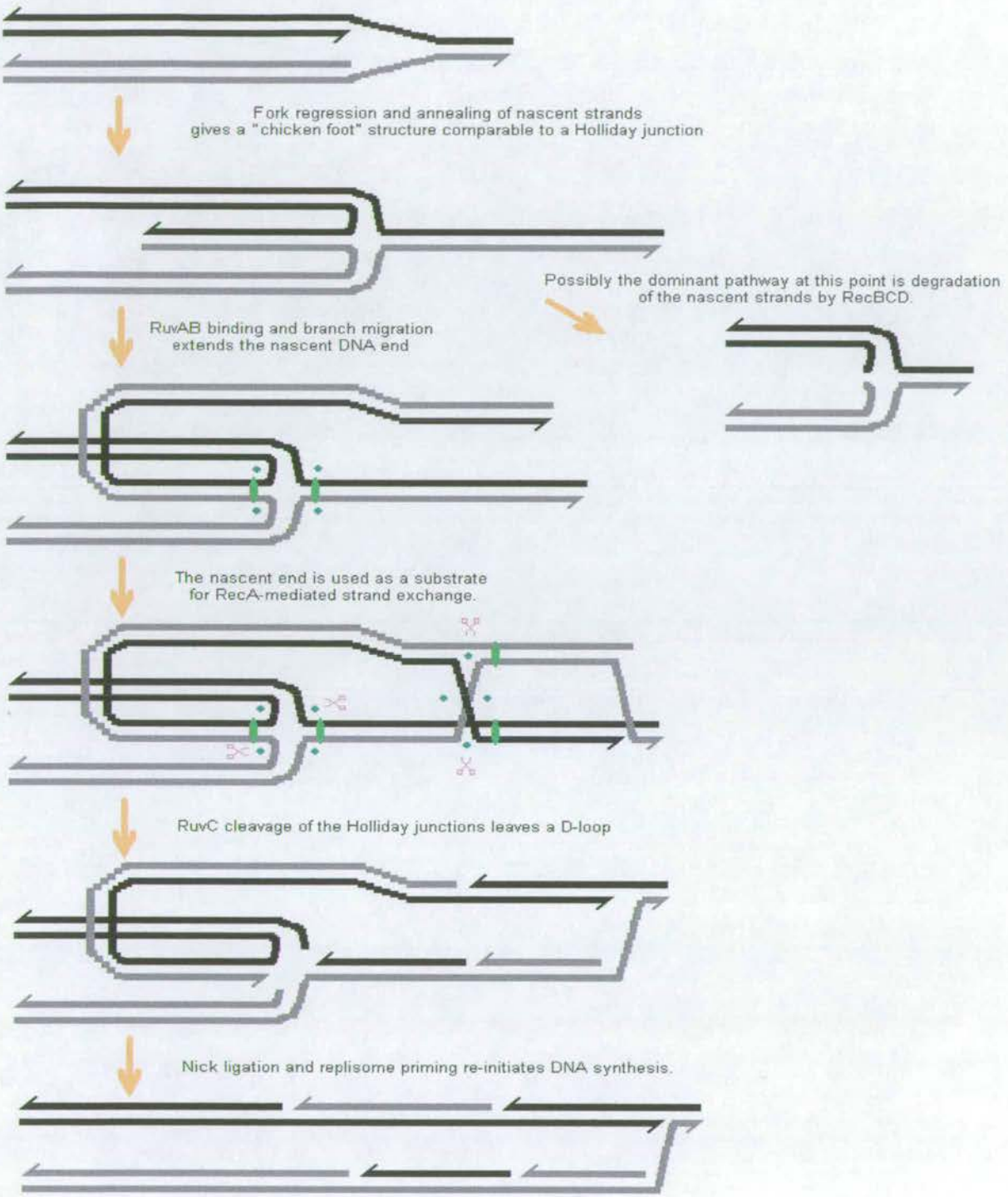
Sensitivity to rich media is observed in *priA*$^-$ cells, probably due to rapid growth entailing less time for restart of stalled replication forks. The processing of these stalled forks into broken forks (substrates for RecA) might also explain why SOS induction is observed in *priA*$^-$ strains (Nurse, Zavitz and Marians, 1991; Kuzminov, 1995b). Crucially, UV repair and conjugational and transductional recombination deficiencies are observed in priA$^-$ strains, providing genetic evidence that replication fork assembly is required for at least some types of recombination. *In vitro* work has shown that PriA is able to recognise D-loop recombination intermediates and initiate primosome assembly at them (Lui, 1999; McGlynn *et al.*, 1997; Nurse, 1999). PriA is also required for inducible and constitutive stable DNA replication (see later), which is one form of stable DNA replication induced by D-loops and R-loops respectively (Masai *et al.*, 1994).

Recombination is needed for replication

Although highly processive, replication forks initiated at *oriC* do not always complete replication of the *E. coli* genome. Forks may be broken or stalled by nucleotide starvation, extremely strong protein-DNA interactions, stalled RNA polymerases during transcription (McGlynn and Lloyd, 2000), or a variety of different mutations and damage to the DNA duplex. After the first few cycles of replication in bacteriophage T4, recombination is required for further rounds of replication initiation (Mosig, 1998). Recombination functions have been proposed to be involved in re-initiation of replication disrupted by UV damage (Rangarajan *et al.,* 2002; Courcelle and Hanawalt, 1999; Courcelle and Hanawalt, 2001). Mutations in replicative functions cause more frequent stalling of the replication fork, are dependent on recombinational activities for viability, and displaying a hyper-rec phenotype (Michel *et al.,* 2000). For example, *rep*$^-$ or *holD*$^-$ strains have reduced fork processivity and are dependent on RecBCD (Flores *et al.,* 2001, and Seigneur *et al.,* 1998). Significantly, *radC*$^-$ strains display increased rates of replication fork stalling, with subsequently elevated levels of recombination in tandem repeats (Saveson and Lovett, 1997).

Stalled replication forks can be re-initiated by fork reversal away from DNA damage, making the lesion accessible to DNA repair enzymes. Reversal of a paused replication fork allows the two template strands to re-anneal, and the single strands of the two nascent strands can also anneal. This produces a "chicken-foot" structure (Higgins *et al.,* 1976) which resembles a Holliday junction (figure 1.12). There is evidence for two methods of fork reversal *in vivo* in the current literature. The first mechanism involves RecA polymerisation across a ss gap formed on one strand as a result of uncoupling of the 2 polymerases in a replication fork. Chicken foot cleavage by RuvABC in a *dnaB*^{ts} *recBC*^{ts} strain is dependent on RecA (Seigneur *et al.,* 1998). The second mechanism was observed in *rep*$^-$ and *holD*^{ts} mutants, and is RecA independent. Here, chicken foot cleavage by RuvABC is detected in *recBC*^{ts} mutants that are *recA*$^-$. This is thought to occur at forks stalled at replication pause sites or bound proteins, where the polymerases are expected to stop almost simultaneously, leaving little or no ss gaps. In both pathways RuvAB is required for the formation of the chicken foot.

Figure 1.12: Stalled fork reversal and processing leads to non-crossover products.



Chicken foot structures or Holliday junctions formed by replication fork reversal can either be resolved by RuvABC (and re-initiated by DSBR), or converted back into three-way junctions by RecBCD-mediated degradation of the paired nascent strands (Michel and Sherratt, 2000). This RecBCD activity at a reversed fork would regenerate a three-way junction suitable for replication fork re-initiation some distance behind the replication blockage, and give the fork a second chance at overcoming the obstacle (figure 1.12).

Recombination is required to regress some stalled forks and reinitiate broken replication forks (Cox, 2002) by using RecA to provide a D-loop. The ds end produced by fork reversal or cleavage of the chicken foot structure can be used by RecBCD-mediated DSBR to generate a D-loop (Flores *et al.*, 2001), from which PriA and other proteins can reinitiate lagging strand synthesis. This is supported by the low viability observed in *recBCD*$^-$, *ruv*$^-$, and especially *priA*$^-$ cells (Seigneur *et al.*, 1998). Broken (otherwise known as "collapsed") forks can occur as replication enters ss gaps or nicks (Kuzminov, 1995a), and are repaired by conventional DSBR leading to re-initiation (see multimerisation chapter 4).

Two other mechanisms of fork reversal have been proposed: spontaneous reversal caused by a build up of positive supercoiling ahead of the replication fork, and RecG-mediated reversal after UV irradiation (McGlynn and Lloyd, 2000). No evidence for the first of these mechanisms has yet been obtained *in vivo* and the argument for the second mechanism remains indirect.

It has been suggested that DNA secondary structures formed in TR tracts might stall replication forks. Repeated stalling and re-initiation by recombination was hypothesised to lead to an "onion skin" structure of multi-branched replication bubbles (Fu *et al.*, 1991), with unequal recombination or replicative strand-switching events leading to large repeat expansions (Kuhl and Caskey, 1993). In keeping with this, blocks to DNA synthesis are known to encourage illegitimate recombination (Bierne and Michel, 1994) through their persistent DNA single stranded regions and ends.

In recent years attention has been cast towards methods of stable DNA replication initiated away from *oriC*, and therefore not subject to the usual tightly controlled method of duplex opening by DnaA. Essentially, two forms of replication priming were observed, Inducible Stable DNA Replication (iSDR), and Constitutive Stable DNA Replication (cSDR). Both require the strand-exchange activity of RecA to insert a polynucleotide primer into a duplex.

Inducible Stable DNA Replication (iSDR) is initiated at recombination intermediates during SOS de-repression

Inducible stable DNA replication (iSDR) occurs during SOS induction. Under these circumstances, replication is thought to be primed on D-loops formed by DSBR. In support of this, iSDR is completely dependent on RecA, RecBC, PriA, and Pol III, with decreased iSDR in *recF*⁻ and *recN*⁻, and stimulated iSDR in *recD*⁻ and *recJ* strains (Asai, Bates, and Kogoma, 1994). Initiation of iSDR in *ruv*⁻ and *recG*⁻ strains is elevated because of persistent D-loops, but the same lack of processing leads to repressed elongation of iSDR, as forks pass through other sites of aborted recombination. The entire *E. coli* chromosome is replicated in a θ manner. Two sites of iSDR initiation (all called *oriM*) are concealed within *oriC*, and one was found in *terC*. Marker frequency experiments indicated at least another two sites elsewhere in the chromosome. These sites are recombination hot-spots, replication is initiated from these sites independently, and choice of site is random. Tus protein is required for this process, presumably to arrest existing forks by blocking the progression of replisome helicase DnaB. An engineered strain of *E. coli* with additional artificially placed *ter* sites requires homologous recombination mediated by RecA, RecB, and RuvABC for viability (Bibnenko *et al.*, 2002). It would seem that replication forks trapped at *ter* sites provide opportunity for recombinational DSBR by RecBCD, with intrinsic initiation of replication (Horiuchi and Fujimura, 1995).

Constitutive stable DNA replication (cSDR) occurs at R-loops in *recG*⁻ and *rnhA*⁻ mutants

A second form of stable DNA replication was found to be constitutive in certain mutants which conferred additional stability to R-loops (RNA transcripts bound to opened DNA duplexes, catalysed by RecA-mediated strand invasion). This cSDR was observed in *rnhA*⁻ cells, in which RNaseHI is no longer present to degrade RNA bound to template DNA (Kogoma *et al.,* 1985a and 1985b). cSDR is also observed in *recG*⁻ mutants, and RecG is a member of the RNA/DNA helicase II superfamily with a characterised ability to disrupt R-loops (Vincent *et al.,* 1996). Since persistent R-loops induce the SOS response, both of these mutants also exhibit iSDR. *rnhA*⁻ *recG*⁻ double mutants are inviable.

At least five sites of transcript-initiated chromosomal replication (*oriK*) were observed in *rnhA*⁻ *Δoric* mutants, including two in the *terC* region of the chromosome. Compared to the *oriC*-DnaA replication system, cSDR is very poorly regulated, being virtually uncoupled to cell mass or cell cycle, and choosing *oriK* sites at random. This has lead to suggestions that it is the remnant of a primitive system of replication used opportunistically by *E. coli* in times of starvation when further protein synthesis is not possible.

DNA replication in pUC18-derived plasmids.

pUC18 is used as a cloning vector for TR tracts in every study presented in this thesis. The method of pUC18 replication initiation and copy number control is of particular relevance to the intra-cell competition experiments carried out in chapter 6. pUC18 replication is unidirectional, allowing clear distinction between leading and lagging strand synthesis.

Plasmid DNA replication is initiated in a transcription-dependent manner. DNA-dependent RNA polymerase generates a several hundred nucleotide transcript from a promoter lying 550 bases upstream of the origin of replication. An R-loop is generated as the transcript forms a stable RNA-DNA hybrid with its template. Endonucleolytic cleavage of the RNA by RNAase H provides a primer (called RNA II) for DNA Polymerase I. PriA binds to this branched DNA substrate, and is stabilised by the co-operative binding of PriB. DnaT then binds to the PriA-PriB complex and recruits PriC, which confers additional stability to the complex. The replicative helicase DnaB is acquired from its competitive complex with DnaB in solution, and completes the "preprimosome" complex. DnaG primase is then able to associate with the preprimosome complex, and lay primers for DNA synthesis, which is carried out after DnaN clamp loading by the γ complex DNA Polymerase III.

Copy number control of pMB1 replicons is dependent on inhibiting the frequency of RNA II pre-primer maturation. A second RNA molecule (called RNA I) is transcribed from the opposite strand in the same region of DNA as RNA II. The two complementary transcripts bind and prevent RNA II adopting a cloverleaf structure necessary for stable R-loop formation (Dooley and Polisky, 1987). In addition, RNA I / RNA II hybridisation is stabilised by a 63 amino acid protein encoded 400 nucleotides downstream of the replication origin by a gene called *rop* (Helmer-Citterich *et al.,* 1988). This system usually generates a copy number of approxiamtely 15-20 plasmids per cell, but pUC plasmids have mutations disrupting the inhibition of priming, allowing them to achieve copy numbers of 500-700 in normal growth conditions. A G to A point mutation near to the site of RNA I initiation causes the RNA I transcript to be truncated by three nucleotides, disrupting the integrity of the 5'

single-stranded domain of RNA I crucial for RNA I / RNA II hybrid formation (Chambers *et al.*, 1988). In addition, the *rop* gene has been removed from pUC plasmids.

The R-loop dependence of plasmid replication initiation means that *recG* or *rnaHI* mutants displaying cSDR would be expected to further increase pUC18 copy number. Also, all of the pre-primosome proteins described above are host-encoded and are long-lived in the cell. In fact, conditions that inhibit protein synthesis (such as the presence of chloramphenicol, or entry into stationary phase) prevent host chromosome replication, but allow plasmids containing the pMB1 replicon to continue replicating until two or three thousand copies have accumulated within each cell (Clewell, 1972).

The SOS response

If a cell encounters highly stressful circumstances when the degree of genetic damage it receives saturates its basal capability for repair, it has the ability to upregulate normal repair capabilities and induce less faithful means of DNA repair/replication in an attempt to maintain viability. A cell whose DNA has been repaired by these more mutagenic pathways may be more compromised in the future, but at least has a fighting chance of survival. This tightly controlled cellular response, which up-regulates several DNA repair pathways, alters DNA replication, and inhibits cell division, is known as the SOS response in *E. coli* (see Bridges, 1998; Humayun, 1998; Smith and Walker, 1998; for reviews). Some of the situations in which *E. coli* activates the SOS response are discussed in Chapter 5 of this thesis. Similar LexA-like regulation of RecA homologues are thought to occur in other gram-negative bacteria (Riera *et al.*, 1994) and a large number of damage-inducible genes have been reported in *Saccharomyces cerevisiae* (Jelinsky and Samson, 1999).

The initial signal for SOS induction is RecA filament-promoted self-cleavage of LexA protein. The LexA binding site is deep in the RecA filament groove, and partly overlaps the DNA2 binding site used to search for homologous sequences. LexA is a 24KDa repressor protein which normally prevents the high expression of several operons of DNA repair genes (Mount *et al.*, 1975; Mount, 1977; Howe and Mount, 1978). LexA auto-cleavage promoted either by persistent recombinogenic ends or by a large amount of recombinogenic DNA within one cell, will cause de-repression of the SOS response. The ssDNA required for RecA loading and subsequent SOS de-repression may be produced by presynaptic activities of recombination (Chaudhury and Smith, 1985), or by attempted replication through damaged DNA (Sassanfar and Roberts, 1990).

Because the affinity of LexA repressor varies from one SOS-response gene promoter to another, significantly increased expression of some SOS response genes requires much less LexA repressor cleavage than others. These "early response" genes usually already have significant gene product present within uninduced cells, and the function of the SOS response seems initially to give the damaged genome more enzymatic activity to repair itself by the conventional means of nucleotide excision repair, and

recombinational repair. Expression of the cell division inhibitor, SfiA, delays chromosome segregation until two complete daughter chromosomes are available for segregation. *dinA/polB* encodes DNA polymerase II which has homologues in humans, yeast, and a selection of viruses and phage (Bonner *et al.*, 1990). However, if these conservative mechanisms fail, the persistent RecA filaments will continue to cleave LexA repressor, leading to the de-repression of "late response" SOS genes. Expression of the UmuDC proteins is very much a desperate attempt to maintain chromosome structure, as they promote a means of error-prone DNA replication called translesion synthesis.

Table 1.3: Genes for which up-regulated expression has been observed as part of the SOS response.

Gene*	basal transcript copies per cell	SOS-induced transcript copies per cell
lexA	1,300	continually degraded
uvrA/dinE	20	250
uvrB	250	1,000
uvrD	5,000-8,000	25,000-65,000
polB/dinA	50	350
dinB	?	?
ruvA	700	5,600
ruvB	200	1,600
dinI	<500	2,300
uvrB	?	?
uvrD	?	?
ftsK	?	?
ssb	?	?
recA	1,000-10,000	100,000
recN (3 copies)	?	10-fold increase
sfiA	?	125-fold increase
umuD	180	2,400
umuC	20	200
cea	?	?
caa	?	?

* early response, moderate response, late response. The consensus LexA binding site sequence is 5'-TACTG(TA)$_5$CAGTA-3' (Lewis *et al.*, 1994). As many as 69 potential genes/operons have been suggested to have LexA binding sites close to promoter regions (Fernandez de Henestrosa *et al.*, 2000). *din* genes have been identified by their *d*amage-*in*ducibile expression.

A cell whose DNA has been repaired by this mutagenic pathway may be compromised in the future, but at least has a chance of survival. This suggests that a population of highly stressed bacteria will exhibit increased diversity. The same can be said of ageing bacterial populations, since the SOS response is induced when cells enter the stationary phase of growth (Taddei *et al.*, 1995).

Positive feedback occurs by increased expression of RecA as part of the SOS response. Negative feedback occurs by LexA acting as a repressor for its own promoter (Brent and Ptashne, 1980), so if the SOS response is successful in DNA repair, the rapid accumulation of uncleaved LexA in the cell prevents further SOS de-repression. Another negative feedback system functions to prevent the formation and reduce the persistence of the RecA-ssDNA filament signal for SOS induction. Din I is an SOS-induced repressor of RecA-mediated homologous pairing able to displace ssDNA from a RecA filament as well as preventing initial RecA-ssDNA binding (Voloshin *et al.*, 2001). The C-terminal of DinI contains a negatively charged α-helix which interacts with RecA by acting as a ssDNA mimic. This also prevents self-cleavage by LexA and UmuD. Since the mutation rate in *E. coli* increases approximately 100-fold after SOS induction, the *umuDC*-encoded polymerase responsible for this undergoes an additional level of regulation. UmuD exhibits sequence homology to LexA around the cleavage site, and must also undergo RecA-mediated self-cleavage for activation (Burckardt *et al.*, 1988; Shinagawa *et al.*, 1988).

However, the penalties inflicted on SOS-induced cells are not restricted to the mutagenic effects of translesion synthesis. The lytic repressor proteins of various labdoid bacteriophages are also cleaved by persistent RecA filaments, but are not as sensitive as LexA (Sauer *et al.*, 1982; Eguchi *et al.*, 1988). Thus repressor auto-cleavage results in prophage induction under circumstances of DNA-damage in which the viability of the cell is greatly reduced (Mount, 1977) or in aging populations of *E. coli* (Taddei, Matic, and Radman, 1995). LexA also functions to repress the expression of the Tn5 transposase gene *tnp* (Kuan *et al.*, 1991). It is interesting to note that expression of lethal proteins from colcinogenic plasmids occurs under conditions of extreme LexA cleavage. Both the induced lytic cycle and expression of Colicin A and E1 would kill the compromised host cell, so have been compared to eukaryotic

apoptosis mechanisms because functionally they prevent inviable cells from wasting finite external resources available to the culture.

Chapter 2: Materials and Methods

Materials

Strains of Bacteria

Table 2-1. Strains of *E. coli* used in this work.

Desig-nation	Genetic Background	Notes	Reference/ Source
DL49	AB1157 *recB$_{21}$ recC$_{22}$ sbcA$_{23}$*	activated ET pathway for recombination	JC8679
DL51	*recB21 recC22 sbcB15*		JC9387
DL68	AB1157 *rec$^+$ su$^-$ F$^-$*	negative control for DL49	JC9937
DL150	AB1157 *recA13*		FWS805 =JC9935
DL324	Wild type JM83 *	Contains λ prophage	JM83
DL474	AB1157 *recB21*		N2362 (Chalker *et al*, 1988)
DL513	wild type AB1157 ≅		N2677 (Chalker *et al*, 1988)
DL515	AB1157 *sbcC201*		N2679 (Chalker *et al*, 1988)
DL654	*recA::Cm^R recD$_{1014}$*		DB1318
DL686	*umuC$^-$*		
DL689	*umuD$^-$*		
DL695	AB1157 *uvrC::Tn10 HCR$^-$ supE44*		
DL696	SY55 *uvrA$_{277}$::Tn10 supE44*		W3110
DL697	AB1884 *uvrC$_{34}$ supE44*		
DL698	AB1885 *uvrB$_5$ supE44*		
DL699	AB1886 *uvrA$_6$ supE44*		
DL887	JM83 *recA::Cm^R*	source of pUC18 in DL324	P1 from DL654 into DL324, (Pan and Leach, 2000).
DL913	DL887 (*recA::Cm^R*)	source of CAG$_{24}$	Cathy Abbott
DL915	DL887 (*recA::Cm^R*)	source of CAG$_{43}$	Cathy Abbott
DL1078	*ΔrecG263::cat*		N4394
DL1083	DH5α	source of pUC4K	DH5α
DL1096	AB1157 *recJ::Tn10* Tc^R		P1 into DL513 (Cromie *et al*, 1999)
Desig-nation	Genetic Background	Notes	Reference/ Source
DL1100	AB1157 *ruvA60::Tn10* Tc^R		P1 into DL513

P1 denotes P1 transduction of an allele from a donor strain to a recipient.

			(Cromie *et al*, 1999)
DL1102	AB1157 *ruvAC65 eda51::Tn10* Tc^R		P1 into DL513
DL1104	AB1157 *recG263::*Km^R		P1 into DL513 (Cromie *et al*, 1999)
DL1108	AB1157 *recO1504::Tn5* Km^R		P1 into DL513 (Cromie *et al*, 1999)
DL1110	AB1157 *recR252::Tn10* Km^R		P1 into DL513 (Cromie *et al*, 1999)
DL1118	AB1157 *ruvAC65 eda51::Tn10* Tc^R		P1 into DL513
DL1121	AB1157 *ruvC53 eda51::Tn10* Tc^R		P1 into DL513 (Cromie *et al*, 1999)
DL1133	*priA2::*Km^R		P1 into DL513 (Cromie *et al*, 1999)
DL1218	*W3110* ΔlacU169 gal490 λcI857Δ(cro-bioA) nadA::Tn10 32°C	32°C working strain for Red/Gam recombination pathway mediated chromosomal engineering	DY329
DL1265	DL887 (*recA::*Cm^R)	source of mouse CTG_{24}	
DL1266	DL887 (*recA::*Cm^R)	source of mouse CAG_{24}	
DL1267	DL887 (*recA::*Cm^R)	source of mouse CTG_{43}	Cathy Abbott
DL1268	DL887 (*recA::*Cm^R)	source of mouse CAG_{43}	Cathy Abbott
DL1275	DL887 (*recA::*Cm^R)	source of synthetic CTG_{28}	
DL1276	DL887 (*recA::*Cm^R)	source of synthetic CTG_{28}	
DL1285	DL1218 *recQ*$^-$$Km^R$	original isolate of *recQ-*Km^R. 32°C	This work
DL1299	DL324 (JM83) *recQ*$^-$$Km^R$	P1 from DL1285	This work
DL1300	DL49 ($recB_{21}$ $recC_{22}$ $sbcA_{23}$) *recQ*$^-$$Km^R$	P1 from DL1285	This work
DL1301	DL513 *recQ*$^-$$Km^R$	P1 from DL1285	This work
DL1310	DL1096 (*recJ::Tn10* Tc^R) *recQ*$^-$$Km^R$	P1 from DL1285	This work
DL1329	DL1218 *yraN*$^-$$Km^R$	original isolate of *yraN*$^-$$Km^R$ .32°C	This work
DL1330	DL49 ($recB_{21}$ $recC_{22}$ $sbcA_{23}$) *yraN*$^-$$Km^R$	P1 from DL1329	This work
DL1331	DL51 (*recB21 recC22 sbcB15*) *yraN*$^-$$Km^R$	P1 from DL1329	This work
DL1333	DL513 *yraN*$^-$$Km^{R-}$	P1 from DL1329	This work
DL1334	DL1100 (*ruvA60::Tn10*) *yraN*$^-$$Km^{R-}$	P1 from DL1329	This work
DL1336	DL1210 (*ΔruvC::*Tc^R *ΔxerC::*Cm^R) *yraN*$^-$$Km^{R-}$	P1 from DL1329	This work
DL1347	FC40 *polA*$^-$ *polB*$^-$		
DL1348	FC40 *polA*$^-$		

Desig-nation	Genetic Background	Notes	Reference/ Source
DL1349	FC40 *wild type*		
DL1350	FC40 *polB*$^-$		PF1360
DL1351	DL1218 *recF*$^-$ *Cm*^{R}	original isolate of *recF*$^-$ *Cm*^{R}.32°C	This work
DL1365	DL324 (JM83) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1368	DL513 *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1369	DL515 (*sbcC201*) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1370	DL1096 (*recJ::Tn10*) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1371	DL1104 (*recG263::Km*^{R}) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1372	DL1106 (*recN262 tyrA16::Tn10* Tc^R) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1373	DL1108 (*recO1504::Tn5* Km^R) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1374	DL1110 (*recR252::Tn10* Km^R) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1376	DL1333 (*yraN*$^-$ *Km*$^{R-}$) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1396	DL1301 (*recQ*$^-$ *Km*^{R}) *recF*$^-$ *Cm*^{R}	P1 from DL1351	This work
DL1397	DL513 *mutS::Tc*^{R}		
DL1399	DL1333 (*yraN*$^-$ *Km*$^{R-}$) *mutS::Tc*^{R}	P1 from DL1397	This work
DL1437	*λ:gal cI*$^+$ *nalR ilv his rpsL*	reporter strain used to detect SOS de-repression by λ:gal assay.	MT101 from Taddei, Matic and Radman, 1995(=FR371)
DL1438	*λ:gal cI(Ind*$^-$*) nalR ilv his rpsL*	control for λ:gal assay	MT105 from Taddei, Matic and Radman, 1995(=FR372)
DL1439	DL1104 (*recG*$_{263}$*::Km*^{R}) *RecJ*$_{284}$*::Tn10*	P1 from DL1096	This work
DL1440	DL1077 (*recG*$_{263}$*::Km*^{R}) *RecJ*$_{284}$*::Tn10*	P1 from DL1096	This work
DL1441	DL1299 (*recQ*$^-$ *Km*^{R}) *RecJ*$_{284}$*::Tn10*	P1 from DL1096	This work
DL1441	DL1299 (*recQ*$^-$ *Km*^{R} in JM83) *recJ::Tn10*	P1 from DL1096	This work
DL1484	DL1078 (*ΔrecG263::cat*) *yraN*$^-$ *Km*$^{R-}$	P1 from DL1333	This work

Desig-nation	Genetic Background	Notes	Reference/ Source
DL1486	*λsfiA.lacZ*	reporter strain used to assay SOS de-repression by β-galactosidase activity.	Millie Masters MG1655 = DY329
DL1510	DL1333 (*yraN*$^{-}$*Km*^{R}) *recA::Cm*^{R}	P1 from DL654	This work
DL1512	DL1334 (*ruvA60::Tn10 yraN*$^{-}$*Km*^{R}) *ΔrecG263::cat*	P1 from DL1078	This work
DL1522	AB1157 *lexA1 (ind*$^{-}$*)*	cannot be induced for SOS response. (SOS^{--})	AB2494
DL1523	AB1157 *lexA3,51 (def), galK2, lac*$^{+}$, *his*$^{+}$, *str*^{R}.	Defective LexA cannot bind to DNA and repress SOS, so is fully induced for SOS response. (SOS^{--})	DM1420
DL1526	*λ:gal cI*$^{+}$ *nalR ilv his rpsL recA1*	control for λ:gal assay	MT112
DL1583	JM83 *umuDC::Cm*^{R}		
DL1554	*recD*$_{1901}$*::Tn10, λ*$^{-}$*, rph-1*	*recD*$_{1901}$*:Tn10* donor strain	CAG12135 from Yale *E. coli* genetic stock center
DL1555	DL1437 *recD*$_{1901}$*::Tn10*	*recD* capable of SOS reporter λ:gal assay	This work
DL1556	DL513 AB1157 *recD*$_{1901}$*::Tn10*	P1 from DL1555	This work

≅ The AB1157 strains are derivatives of: (Howards- Flanders and Theriot, 1966), F^{-} hisG4 argE3 Δ (gpt- proA2)62 thr-1 thi-1 leuB6 kdgK51 rfbD1 mgl-51 ara-14 lacY1 galK2 xyl-5 mtl-1 tsx-33 supE44 rpsL31 (Str^{r}).

* The JM83 strains are derivatives of: *ara Δ(lac-proAB), rpsL,* ϕ80d*lacZ*ΔM15 (Yanisch-Perron *et al,* 1985)

Plasmids

Table 2-2. Plasmids used in this work.

Plasmid	Description	Notes	Reference/ Source
pUC18	2686bp Amp^R	500-700 copy number	
pUC4K	3966bp Km^R	500-700 copy number	Pharmacia LKB Biotechnology
pACYC184	4000bp Cm^R Tc^R	10 copy number	
CAG_{43}	mouse $(CAG)_{43}$ in pUC18 EcoR1 site		Cathy Abbott
CTG_{43}	mouse $(CTG)_{43}$ in pUC18 EcoR1 site		Cathy Abbott
GCC_{24}	synthetic $(GCC)_{24}$ in pUC18 EcoR1 site		Cathy Abbott
CGG_{24}	synthetic $(CGG)_{24}$ in pUC18 EcoR1 site		Cathy Abbott
CAG_{28}	synthetic $(CAG)_{28}$ in pUC18 EcoR1 site		
CTG_{28}	synthetic $(CTG)_{28}$ in pUC18 EcoR1 site		
CAG_{25}	mouse $(CAG)_{25}$ in pUC18 EcoR1 site		
CTG_{25}	mouse $(CTG)_{25}$ in pUC18 EcoR1 site		
TA	pLDR11 *Tet^R, Amp^R*	Maintained in DL1139	
TAR	$(CTG)_{39\text{-}43}$ in pLDR11 EcoRI site	Maintained in DL1181	Camelia Miheascu
TARi	$(CAG)_{39\text{-}43}$.in pLDR11 EcoRI site	Maintained in DL1180	Camelia Miheascu

(All trinucleotide repeat orientations given are on the lagging strand during plasmid replication.) Purity of repeats was confirmed by sequencing.

Synthetic Oligonucleotides

Table 2-3: Synthetic oligonucleotides for PCR and primer extension reactions.

Identification	Sequence (5'→3')	Notes	Reference/ Source
M6833	CCCCTTTCTAGC CTTCTTCA	primer for flanking sequence of cloned $(CTG/CAG)_{43}$ repeat	Abbott and Chambers (1994)
M6834	TTTGGTCCAAAC GGGATGCT	opposite primer for flanking sequence of cloned $(CTG/CAG)_{43}$ repeat	Abbott and Chambers (1994)
ORIENTN	CCTCTTCGCTAT TACGCCAG	primer on pUC18 vector used in combination with either M6833 or M684 to determine repeat length and orientation simultaneously	
FP1	**GCAGTTTCTCG CGCAGGCGCTG AAAATAGCGCC TGTTTTTATTTC AGGCAATCGGG GTGA**AAGCCAC GTTGTGTCTCAAA	***recQ* upstream flanking sequence** fused to kanamycin resistance gene priming sequence	This work
FP2	**CATCAACATACA TTGACTCGCGGG GGAAACGTTTAC GGAGTCTTCATA CTGGCACTTTT**TC GTGAAGAAGGTGT TGCTG	**reverse complement of *recQ* downstream flanking sequence** fused to kanamycin resistance gene complement strand priming sequence	This work
SQ1	GTGGTTGGCAAAT CTGGAAT	diagnostic primer for *recQ* locus	This work
SQ2	ATGGCGACGGTGA AATAA	diagnostic primer for *recQ* locus	This work

DNA sequence information was obtained from Entrez nucleotides (http://www3.ncbi.nlm.gov/Entrez/index.html). Primers were chosen using Gene Jockey™ and Primers 3™. All oligonucleotides designed in the course of this work were synthesised by OSWEL DNA Service UK (www.oswel.com).

Identification	Sequence (5'→3')	Notes	Reference/ Source
W1	**ATGGCTACAGTAC CAACAAGGTCAGG TAGTCCCCGTCAG TTAACCACCAAAC AGACCGGC**AAGCC ACGTTGTGTCTCAA A	***yraN* upstream flanking sequence** fused to kanamycin resistance gene priming sequence	This work
W2	**TTATGAGTGGTCA TTAAAGGCATCCT TAATCCACTCAAC CTCATTCCCGGTG AAGGCTAC**TCGTG AAGAAGGTGTTGCT G	**reverse complement of *yraN* downstream flanking sequence** fused to kanamycin resistance gene complement strand priming sequence	This work
WH1	ACGGGTAGCCCGA TTACTCT	Diagnostic primer to screen the *yraN* locus.	This work
WH2	TGTTGCCATTGAGC AGAGAC	Opposite diagnostic primer to screen the *yraN* locus.	This work
recF KO1	**GCCAGAGCGCGG CTTATGTTGTCAT GCCAATGAGACTG TA**TGTGACGGAAG ATCACTTCG	***recF* upstream flanking sequence** fused to chloramphenicol resistance gene priming sequence	This work
recF KO2	**ATTCGACATCAAC GTTTCTCGCTCAT TTATACTTGGGTT A**CTGCCATTCATCC GCTTATT	**reverse complement of *recF* downstream flanking sequence** fused to chloramphenicol resistance gene complement strand priming sequence	This work
recF test1	GCGGCTTATGTTGT CATGC	diagnostic primer for *recF* locus	This work
recF test2	TTCGACATCAACGT TTCTCG	opposite diagnostic primer for *recF* locus	This work

Media

BBL Agar

10g Trypticase (Baltimore Biological Laboratories), 5g NaCl, 10g Bacto-agar (Difco) per litre, adjusted to pH 7.2 with NaOH

BBL Top Agar

Same as BBL agar, but made with 6.5g Bact-agar (Difco) per litre.

LB Agar

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 15g Bacto-agar (Difco) per litre, adjusted to pH 7.2 with NaOH.

L Broth

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl per litre, adjusted to pH 7.2 with NaOH.

L broth P/C

98ml LB, 1ml 20% maltose, 0.5ml $MgSO_4$.

Lc Agar

10g Tryptone, 5g yeast extract, 5g NaCl and 10g Difco-agar per litre. The pH was adjusted to 7.2 using NaOH.

Lc Top Agar

As Lc agar, but containing 7g Difco-agar.

MacConkey Agar

Supplied by Oxoid. Xg per litre of sterile water.

20g Peptone, 10g lactose, 1.5g bile salts n°3, 5g NaCl, 0.03g Neutral red, 0.001g crystal violet, 15g agar, pH 7.1. The mixture was heated to boiling point for only the length of time required for the solution to become transparent. The medium was not

autoclaved (because of the production of toxic by-products which would stress the *E. coli*), but was surface sterilised with a UV lamp after plate pouring.

Minimal Agar plates

15g Difco agar per litre sterile water. 10.5g K_2HPO_4, 4.5g $KHPO_4$, 1g $(NH_4)_2SO_4$, 0.5g sodium citrate.$2H_2O$, 1ml from 1M $MgSO_4$ stock, 0.5ml from 1M stock vitamin B1 thiamine hydrochloride (excess), 4ml from a $10mgml^{-1}$ stock of casamino acids (excess), 10ml from a 20% stock solution of glucose, antibiotics as required.

Glucose minimal liquid medium

0.75l sterile distilled water and 250ml Spitzizen salts were supplemented with glucose (to 0.2% by volume), and $15\mu g\ ml^{-1}$ of threonine, histidine, arginine, and leucine.

SOC

SOC was prepared from 100ml of L-broth with additional glucose (20mM final concentration), $MgCl_2$ (10mM), and $MgSO_4$ (10mM).

Spitzizen salts

10g $(NH_4)_2SO_4$, 70g K_2HPO_4, 30g KH_2PO_4, 5g sodium citrate ($Na_3C_6H_5O_7$) and 1g $MgSO_4$ per litre.

Media Additives

Ampicillin ($100mg\ ml^{-1}$ stock)

Ampicillin (Beecham Pharmaceuticals) was dissolved in sterile water and stored at -20°C. It was used at a concentration of $100\mu g\ ml^{-1}$.

500mM $CaCl_2$ stock

Dissolved in sterile, distilled water, autoclaved.

casamino acids

$10mgml^{-1}$ stock. Sterile filtered.

Chloramphenicol ($20mg\ ml^{-1}$ stock)

Chloramphenicol (Sigma Chemical Company) was made up in 100% ethanol and used at 50 $\mu g\ ml^{-1}$. It was stored at -20°C.

20% (w/v) Glucose Stock
Dissolved in distilled water, filter sterilised.

Kanamycin (50mg ml^{-1} stock)
Kanamycin was dissolved in distilled, sterile water and used at 50$\mu g\ ml^{-1}$. It was stored at -20°C.

1M $MgSO_4$
Dissolved in sterile, distilled water, autoclaved.

Mitomycin C
MMC (2mg vial from Sigma-Aldrich) 1$\mu g\mu l^{-1}$ stock in sterile water. Frozen in a clearly labelled box in 0.5ml aliquots. Always stored and used in the dark.

Tetracycline (15mg ml^{-1} stock)
Tetracycline was dissolved in 50% ethanol, 50% sterile water. The tube was wrapped in foil and stored at -20°C. Tetracycline was used at a concentration of 15$\mu g\ ml^{-1}$.

Streptomycin (100mg ml^{-1} stock)
Streptomycin was dissolved in sterile water, stored at –20°C, and used at a concentration of 100$\mu g\ ml^{-1}$.

vitamin B1 thiamine hydrochloride stock
Made to 1M concentration, then sterile filtered.

General Stock Solutions

BSA (Bovine Serum Albumin)

BSA (New England Biolabs) was supplied at a concentration of 20mg ml^{-1} and stored at -20°C.

$CaCl_2$ (100mM)

Dissolved in distilled water, autoclaved

500mM EDTA Stock

500mM EDTA (Sigma Chemical Company), adjusted to pH 8 using glacial acetic acid, autoclaved.

Ethidium-Bromide DNA staining solution Solution (50X stock)

1% (w/v) ethidium bromide in sterile, distilled water

Magnesium Chloride (100mM $MgCl_2$)

MOPS-Glycerol

100mM MOPS [3-(4-morpholinyl) 1-propanesulfonic acid]- NaOH (Sigma Chemical Company, pH 6.5), 50mM $CaCl_2$, 20% (v/v) glycerol.

ONPG solution.

ONPG (O-Nitrophenyl-β-D-Galactopyranoside) was dissolved in Z-buffer at a concentration of 4mg/ml. Prepared fresh each time.

Phage Buffer

3g KH_2PO_4 ,7g HPO_4 , 5g NaCl, 1mM $MgSO_4$, 1mM $CaCl_2$,1% (w/v) gelatine.

3M Sodium Acetate (pH 5.3)

0.19 volumes of sterile 3M acetic acid were added to 0.81 volumes of sterile 3M sodium acetate, and autoclaved.

10 x TBE (Tris- Borate-EDTA-Buffer) pH 8

54g Tris-base, 27.5g Boric acid (Fisher Scientific), 20 ml of 500mM EDTA per litre.

TM buffer

10mM Tris, 10mM$MgSO_4$.

10 x Tris-EDTA buffer

100mM Tris-base (Sigma Chemical Company), 10mM EDTA, adjusted to pH 7.5 with concentrated NaOH, autoclaved.

1M Tris-HCl Stock

1M Tris base, adjusted to pH 7.5 using concentrated HCl, then autoclaved.

Vistra Green DNA staining stock solution in DMSO (Amersham Pharmacia Biotech) Divided into 50µl aliquots and the tubes wrapped in silver foil, stored at -20°C. Dilute 1:10,000 in fresh gel buffer for 45 minute staining incubations.

VT fixing solution for polyacrylamide gels

2Litres composed of 200ml methanol, 200ml acetic acid, 1600ml dist. water.

Z buffer (100mM phosphate).

For 1l: 16.1g $Na_2HPO_4.7H_2O$ (60mM), 5.5g $NaH_2PO_4.H_2O$ (40mM), 0.75g KCl (10mM), 0.246g $MgSO_4.7H_2O$ (1mM), 2.7ml β-Mercaptoethanol (50mM). pH7. Store at 4°C.

Qiagen Solutions for extraction of plasmid DNA from *E. coli*

The following solutions supplied in the QIAGEN Plasmid Maxi and QIAGEN Plasmid Mini Spin Kit were used for the preparation of plasmid DNA from 100ml and 5ml overnight cultures of *E. coli*.

Resuspension Buffer P1

50mM Tris-HCl (pH 8.0), 10mM EDTA, 100μg ml^{-1} RNaseA, stored at 4°C.

Lysis Buffer P2

200mM NaOH, 1% (w/v) SDS (sodium dodecyl sulphate), stored at room temperature.

Neutralisation Buffer N3

3M potassium acetate, pH 5.5

Equilibration Buffer PB

750 mM NaCl, 50mM MOPS (pH 7.0), 15% (v/v) ethanol.

Elution Buffer EB

1.25 M NaCl, 50mM Tris-HCl (pH 8.5), 15% (v/v) ethanol.

Solutions for Agarose and Polyacrylamide Gel Electrophoresis

TAE agarose gel electrophoresis buffer (10 X)

242g Tris, 100ml 0.5M Na_2EDTA (pH 8.0) and 57.1ml glacial acetic acid.

TBE agarose gel electrophoresis buffer (10 X)

0.9M Tris-borate, 20mM EDTA (pH 8.0)

Agarose Gel Loading Buffer (5X)

20% (v/v) sterile glycerol, 0.05% (w/v) bromophenol blue

5% Gel Solution for Native Polyacrylamide Gels (1.2x TBE)

8.4ml of 10x TBE buffer, 7ml of 50% polyacrylamide gel solution (Flowgen), 54ml of distilled water.

0.6xTBE Gel Running Buffer
60ml of 10x TBE buffer stock solution, 940ml of distilled water

10% (w/v) AMPS
10% (w/v) AMPS (ammonium persulphate; Sigma Chemical Company) was freshly prepared in distilled water.

TEMED
TEMED (N-N-N'-N'-tetra-methyl-1,2-diamino-ethane) (Sigma Chemical Company) was stored at 4°C.

Polyacrylamide Gel Loading Buffer (10x)
95% Formamide, 20mM EDTA, 0.05% (w/v) bromophenol blue.

Restriction Endonucleases and Incubation Buffers

All restriction endonucleases, listed in Table 2-2, were incubated with the buffers provided by the suppliers at the temperatures recommended in the manufacturer's instructions. Bovine serum albumin was added (to a final concentration of $100\mu gml^{-1}$) to stabilise restriction endonucleases over long incubations.

Table 2-4. Restriction endonucleases used in this work.

Enzyme	Cleavage site (5'->3')	Supplier
*Eco*RI	G↓AATTC	New England Biolabs
*Sau*3A	↓GATC	Roche
*Pvu*II	CAG↓CTG	Roche
*Dpn*I	$G^{ME}A$↓TC	Roche
*Bam*H1	G↓GATCC	Roche

Materials for Polymerase Chain Reactions

Taq DNA Polymerase (5U μl^{-1})

Taq DNA Polymerase (Roche) was used with 1 x PCR reaction buffer as supplied by the manufacturer. Enzyme and buffer were stored at -20°C.

dNTP-4 Stock Solution (10x)

dNTP-4 Stock Solution was a mixture of dATP, dTTP, dGTP and dCTP (Roche), each at a concentration of 2 mM, prepared with sterile distilled water and stored at -20°C.

Stratagene Opti-Prime™ PCR Optimisation kit

The twelve reaction buffers in this optimisation kit vary in the concentration of $MgCl_2$, KCl, and pH as follows:

10 mM Tris-HCl	$MgCl_2$	25mM KCl	75mMKCl
pH 8.3	1.5 mM	buffer 1	buffer 2
pH 8.3	3.5 mM	buffer 3	buffer 4
pH 8.8	1.5 mM	buffer 5	buffer 6
pH 8.8	3.5 mM	buffer 7	buffer 8
pH 9.2	1.5 mM	buffer 9	buffer 10
pH 9.2	3.5 mM	buffer 11	buffer 12

Once the correct reaction buffer has been identified, any of six adjuncts may be added to the reaction mix to further optimise the PCR reaction:

Bovine Serum Albumin (BSA) used at 10-100µg/ml.

Formamide used at 1.25-10% (v/v).

Dimethylsulfoxide used at 1-10% (v/v).

Glycerol used at 15-20% (v/v).

$(NH_4)_2SO_4$ used at 15-30mM.

Perfect Match® DNA polymerase enhancer used at 0.1-1U100μl^{-1}.

Materials for end-labelling of DNA

DNA Polymerase I Large (Klenow) Fragment

In this work, Klenow enzyme [supplied concentration: 2U μl^{-1} (Roche)] was used after incubation of plasmid DNA with EcoRI restriction endonuclease (20U μl^{-1}) in EcoRI buffer (50mM NaCl, 100mM Tris-HCl, 10mMMgCl_2, 0.025% (v/v) Triton X-100, pH 7.5). Klenow enzyme was added to unpurified reaction mixtures containing EcoRI and 1 x EcoRI restriction endonuclease buffer.

Radionucleotide *[α-^{35}S] dATP*

[α-^{35}S] dATP was supplied by Amersham Pharmacia Biotech at a concentration of 10mCiml^{-1} in 5 mM Tris-HCl (pH 7.4-7.5). Store at 4°C and use within 8 weeks for best results (although ^{35}S has a half life of 87.4 days, the biological activity of dATP degrades rapidly).

Methods

Bacterial Culture Methods

Storage of Bacteria

Glycerol stocks of *E. coli* strains were prepared in 1.5-ml Eppendorf tubes by adding 0.5 ml of sterile 100 % glycerol to 1 ml of a stationary phase bacterial culture. The tube was sealed with parafilm, labelled and stored at - 70°C.

Dry agar stabs of *E. coli* strains serve as a back-up to the glycerol stocks. They were made from a purified colony that was picked with a sterile wire loop from a freshly streaked LB agar plate. The colony was transferred to a small 2ml plastic tube filled with LB agar, sealed with parafilm and incubated overnight at 37°C. This tube was then stored at room temperature.

Growth of Overnight Cultures

To make overnight cultures of *E. coli* the desired strain was streaked on LB agar plates from a frozen glycerol stock to obtain single colonies. A bacterial culture was then grown overnight by inoculating 5 ml of L broth with a single colony from the LB agar plate and shaking at 37°C. Appropriate antibiotics were used to select the correct strain at all stages.

Mitomycin C incubations and safe disposal

Mitomycin C incubations were carried out in the dark for 25 min.
For the safe disposal of mitomycin C (not exceeding $0.5 mg ml^{-1}$), the following rapid protocol was used. 1.5ml of 5.25% fresh NaOCl solution was added to each ml of MMC solution. After standing for 10 min, any excess NaOCl was removed by adding 1% sodium bisulphite solution. The reaction was then neutralised to a pH between 5 and 9, and disposed of down the drain.

β-galactosidase assay

The cells in each Z-buffer assay mixture were lysed by the addition of 2 drops of chloroform and 1 drop of 0.1%SDS, followed by vortexing for 10 seconds. The tubes

were incubated at 28°C for 5 min. 0.2ml of ONPG solution was added, and the reaction mixture pipetted into a pre-calibrated 1cm plastic cuvette. The cuvette was placed in a spectrophotometer set to take a "time-scan" measuring the appearance of product at 420nm over 10 min.

Transformation of Bacterial cells

Preparation and Storage of $CaCl_2$-Competent *E. coli* cells

Competent cells were made by diluting an overnight culture of the appropriate *E. coli* strain 10-fold in 20ml of L broth. The culture was grown by shaking at 37°C until OD_{650}=0.4 to 0.5 was reached. The cell culture was incubated on ice for 20 min and then centrifuged at 5 krpm for 5 min at 4°C (Centra-3, International Equipment Company, UK). The supernatant was discarded and the cell pellet was resuspended in 5 ml of ice-cold 100mM $MgCl_2$. This step was followed by centrifugation at 5 krpm for 5 min at 4°C and the supernatant was removed. The cell pellet was resuspended in 5 ml of ice-cold 100mM $CaCl_2$ and incubated on ice for 20 min. After centrifugation at 5 krpm for 5 min at 4°C the supernatant was discarded and the cell pellet resuspended in 400 µl of MOPS-glycerol solution. Aliquots of 200 µl were dispensed into pre-cooled 1.5ml Eppendorf tubes. Tubes were immediately transferred to -70°C.

Transformation of $CaCl_2$-Competent *E. coli* Cells

$CaCl_2$-competent cells of the appropriate *E. coli* strain were thawed on ice. 100ng (1 µl) of plasmid DNA were added to 100 µl of $CaCl_2$-competent cells and gently mixed by pipetting. After at least 20 min of incubation on ice the DNA/cell mixture was heat-shocked at 42°C for 90 seconds. It was then put on ice for 2 min. 900µl L broth was added, and the cells incubated at 37°C for one hour. 250µl (or less for more competent cell preparations) of the transformation mixture was plated on LB agar plates supplemented with the appropriate antibiotics to select for transformants. The plates were incubated at 37°C overnight.

Preparation of Electrocompetent *E. coli.*

1ml of overnight culture was inoculated into 50ml LB (with selective antibiotic) and grown in a 500ml flask at 37°C and 200 revolutions per minute until the OD_{650} was between 0.4 and 0.5. The cultures were then transferred to 25ml centrifuge tubes, cooled on ice for 15 min, and spun in a pre-cooled centrifuge at 7,000rpm (5,500g), 4°C, for 8 min. After discarding the supernatant, the pellets were washed three times in cold, sterile milliQ water (Millipore Mili-Q^{50}) more to reduce the concentration of salt ions capable of producing sparking during electroporation. The final pellets were resuspended in 150µl sterile milliQ water with 20% glycerol, divided into 20µl aliquots, and immediately snap frozen in liquid nitrogen before storage at -70°C. All solutions and containers through out this protocol were sterile and allowed to chill to 4°C before contact with the cells.

Transformation of electrocompetent *E. coli*

1-2µl of DNA in TE was added to 40µl of electrocompetent cells, and incubated on ice for 10 min. Electroporation (Dower *et al.* was performed in ice chilled 0.1cm aperture cuvettes (Invitrogen) on a Bio-Rad gene pulser at 1.8kV, 25µF, and 200ohms. A successful time interval for transformation was in the region of 0.6ms. Immediately after pulsing, 1ml SOC L-broth at 37°C was added, and the entire contents of the cuvette removed to a warm 1.5ml centrifuge tube for 1 hour at 37°C, before plating out 250µl on selective medium.

Transfer of mutations between *E. coli* strains by P1 transduction

Upon infection of *E. coli* cells, some P1 phage particles package random fragments of the bacterial chromosome (up to 2 min), which can then be injected into the recipient *E. coli* strain (for reviews see Ikeda and Tomizawa, 1965a,b,c; Watanabe, 1968; Newman and Masters, 1980). Hence, the transfer of a mutation of interest from one *E. coli* strain to another can be achieved by P1 transduction (J.H.Miller, A short course in bacterial genetics). For that purpose, a P1 lysate was grown on a *E. coli* strain that carries the mutation of interest linked to a cotransducible selective marker, such as a Tn*10* or Tn*5* insertion. The recipient *E. coli* strain is then transduced with the P1

lysate, which provides a substrate for gene conversion on the recipient chromosome. Transductants are identified by selection for the cotransduced marker. The presence of the mutation of interest is then confirmed by testing for the mutant phenotype or diagnostic PCR of the allele.

Preparation of P1 Plate Lysate

A fresh overnight culture of the appropriate *E. coli* strain was diluted 10-fold in fresh L broth supplemented with 2.5 mM $CaCl_2$. This culture was grown for 2 hours at 37°C with shaking. Of this culture 200 µl were added to 100 µl of P1 lysate (10^7 pfu ml^{-1}). After incubation at 37°C for 30 min 2.5 ml of LC top agar containing 5 mM $CaCl_2$ were added to the phage/ bacteria mixture and poured onto a fresh LC agar plate, also supplemented with 5 mM $CaCl_2$. After incubation at 37°C for 6 to 8 hours, 5 ml of phage buffer were applied onto the plate. After 5 miutes soaking time, this phage buffer, together with the top agar was scraped into a 30ml glass bottle containing 100 µl of chloroform. The mixture was vortexed and then incubated overnight at 4°C. After centrifugation at 5 krpm for 10 min the clear supernatant was transferred to a 5ml McCartney bottle (previously rinsed with water and sterilised to remove any detergents) containing 200 µl of chloroform. The McCartney bottle was wrapped in aluminium foil to exclude light, and the P1 lysate was stored at 4°C.

Storage of bacteriophage

Both P1 and λ phage were stored in the dark at 4°C in glass 1/2oz bijou bottles. The bottles had been thoroughly rinsed prior to autoclaving to remove any traces of detergent.

Measuring the titre of a P1 lysate

For the problematic construction of low viability strains containing two or more mutations in separate DNA repair pathways, it was important to verify the quality of the P1 lysate by performing a titre with wild type cells. This was done by diluting the lysate to obtain a useful range of 10^{-4} to 10^{-9}, then applying 100µl of each of these dilutions onto an Lc plate supporting a mixture of *E. coli* and Lc top agar. The lysate solutions were allowed to dry before the plates were incubated at 37°C for six to ten hours. Subsequently, plates containing separate and distinct plaques were chosen to

count the number of plaque forming units present in that dilution. This process was repeated over several plates, and an average pfu value taken.

P1 Transduction

An overnight culture of the recipient *E. coli* strain was grown in 5 ml of L broth supplemented with 2.5 mM $CaCl_2$. 1ml aliquots of this culture were transferred to 1.5ml Eppendorf tubes and cell pellets were obtained by centrifugation at 9krpm for 5 min. The supernatant was removed from each tube and each cell pellet was resuspended in 100 µl of L broth containing 2.5 mM $CaCl_2$. To the first tube 100 µl of undiluted P1 lysate were added while 100 µl of a 10-fold dilution of the P1 lysate were added to the second tube. 100 µl of phage buffer were added to the third tube which served as a negative control. A fourth tube contained 100 µl of undiluted P1 lysate, but no recipient. All four tubes were incubated at 37°C for 20 min. Then 800 µl of L broth supplemented with 2 mM sodium citrate (to stop P1 infection) were added to each tube and incubation was continued at 37°C for 60 min. 250µl of all four transduction mixtures were plated on LB agar plates containing the appropriate antibiotic for selection of transductants.

Plating of recipient cells (tube 3) and of P1 lysate (tube 4) alone should not yield any colonies that are resistant to the appropriate antibiotic. Only if this was the case, single antibiotic-resistant colonies that had been obtained by plating transduction mixtures from tubes 1 and 2 were purified and tested for the mutant phenotype.

Bacteriophage Lambda Protocols

Harvesting bacteriophage Lambda from plate lysates

250µl of a plating culture of *E. coli* was mixed with a range of ten fold dilutions of a phage λ lysate in 5ml test tubes. After a 10 min incubation at room temperature, 2ml of cooling (46°C) molten LC top agar was added to each test tube. After mixing, the contents of each tube were poured onto fresh LC plates supplemented with 3% (w/v) maltose, 10µg ml^{-1} vitaminB1, 0.04mM $FeCl_3$ and 8mM $CaCl_2$. These plates were incubated, lid upwards, for at six to ten hours at 37°C. After this time, plates giving near-confluent lysis (plaques almost touching) were selected. 5ml TM buffer was poured onto these plates, and the top agar macerated with a sterile glass pipette tip

held at an obtuse angle. The top agar/ TM buffer mixture was carefully scraped into a 30ml McCartney bottle, and 100µl of chloroform added to kill any remaining bateria. The mixture was vortexed and left overnight at 4°C to allow phage diffusion into the TM buffer. The top agar was removed the following day by a ten minute centrifugation at 9,000g and the supernatant transferred to a sterile 1/2oz bijou bottle containing 100µl of chloroform.

Lambda phage spot test

An overnight culture of the *E. coli* strain to be tested was diluted ten fold in L broth P/C and grown for 2.5h at 37°C. An equal volume of TM buffer was added. 250µl of this mixture was then added to 3ml cooling melted BBL top agar (46°C), and poured onto BBL agar plates. Once set and dried, 10µl spots of serial dilutions of lambda were applied to the surface. After overnight incubation the plates were examined for plaques indicating lambda growth.

Methods of DNA Purification

Phenol/Chloroform purification of DNA

Initially, phenol chloroform purification was used to remove a crude DNA extract from a solution also containing protein. For this, an equal volume of phenol was added to the sample in a microfuge tube. The tube was vortexed until an emulsion formed. The tube was centrifuged at 12,000g for 3 min at room temperature. A pipette was then used to carefully transfer the upper aqueous phase to a new tube. An equal volume of chloroform was added, vortexed, and centrifuged as before. The subsequent aqueous phase was removed carefully, so as not to carry over any chloroform, or any of the protein visible at the interface.

DNA Precipitation

Since DNA pellets resulting from an isopropanol precipitation are transparent, pellet paint was used. 2µl of pellet paint (Novagen) was added to the nucleic acid sample (regardless of volume) followed by 0.1 volumes of 3M Na Acetate at room temperature. After brief mixing, 2.5 volumes of ethanol were added, followed by more vortexing and a 2 minute incubation at room temperature. The sample was spun

at 14,000-16,000g for 5 min to precipitate the DNA as a pink pellet. The precipitate was washed with 70 % (v/v) ethanol, air-dried, and resuspended in the appropriate volume of 10 mM Tris-HCl buffer and stored at -20°C. The pellet paint did not have any unwanted downstream effects in subsequent protocols.

Extraction of Plasmid DNA from *E. coli* using Qiagen Mini-prep and Maxi-prep kits

Principle of the Procedure

All plasmid DNA was prepared using the solutions provided by the range of QIAprep Kits (Qiagen Inc.) and by following the manufacturer's instructions. The protocol is an application of a modified alkaline lysis method by Birnboim and Doly (1979) followed by binding of the DNA to a silica-gel membrane in the presence of high salt (Vogelstein and Gillespie, 1979).

Small Scale Preparation of Plasmid DNA (mini-prep)

Small amounts of plasmid DNA (10-15 μg) were prepared from 5ml overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen). For this purpose, cells from 4.5 ml (3 x 1.5 ml) of a fresh overnight culture were pelleted in a single 1.5 ml Eppendorf tube by centrifugation at 15 krpm for 1 minute in a bench top centrifuge (Sorvall Microspin 24). This and all subsequent steps were carried out at room temperature. The cell pellet was resuspended in 250 μl of buffer P1 using a vortex. Bacteria were lysed by adding 250 μl of buffer P2. The tube was inverted gently five times to achieve sufficient mixing of the solutions without shearing the bacterial chromosomal DNA. Addition of 350 μl of buffer P3 neutralises and adjusts to high-salt binding conditions in one step. The solution was centrifuged for 10 min at 15 krpm and the clear supernatant was loaded onto a QIAprep spin column. After another round of centrifugation (1 minute, 15 krpm) the flow-through was discarded and 500 μl of wash buffer PB were added to the column to remove trace nuclease activity and carbohydrate. The column was washed again by adding 750 μl of buffer PE and centrifuging for 1 minute at 15 krpm. This centrifugation step was repeated after the flow-through had been discarded from the collection tube, so that all traces of ethanol present in buffer PE were removed. The column was placed in a sterile 1.5-ml

Eppendorf tube and the DNA released from the membrane by allowing a 1 minute incubation in 100 µl of 10 mM Tris-buffer (warmed to 50°C), before eluting from the column by centrifuging at 15 krpm for 1 minute. The DNA was stored at -20°C.

Large Scale Preparation of Plasmid DNA (Maxi-prep)

A single colony picked from a freshly streaked LB agar plate (supplemented with the appropriate antibiotic) was inoculated into 100 ml L broth containing the antibiotic drug as before. The culture was grown with vigorous shaking overnight at 37°C. The cells were collected by centrifugation at 12 krpm for 6 min at 4°C (Sorvall centrifuge, GSA or SS34 rotor). All subsequent centrifugations were carried out at 16 krpm and 4°C (Sorvall centrifuge, GSA or SS34 rotor).

The cell pellet was resuspended in 4 ml of buffer P1. The cells were lysed in the presence of RNase (provided in buffer P1) by adding 4 ml of buffer P2. The viscous solution was mixed gently by inverting the tube a few times. Bacterial lysis was allowed at room temperature for 5 min. The solution was neutralised and lysis stopped with 4 ml of chilled buffer P3. The samples were mixed and incubated on ice for 15 min. The mixture was centrifuged for 30 min to separate cell debris and chromosomal DNA from the clear solution containing the plasmid DNA. The supernatant was carefully decanted and applied to a QIAGEN-tip 500, which had been equilibrated with 4 ml of QBT buffer. The column was allowed to empty by gravity flow. Then the column was washed twice by permitting 2 x 10 ml of buffer QC to passthrough the column by gravity flow. The DNA was eluted from the column with 5 ml of buffer QF. The DNA was precipitated with 0.7 volumes of isopropanol and centrifugation for 30 min. The supernatant was removed and 2 ml of ice-cold 70 % (v/v) ethanol were added to wash the DNA pellet. A final centrifugation was carried out for 10 min, all liquid was removed from the tube and the pellet was air-dried. DNA was resuspended in 500 µl of 10 mM Tris-HCl buffer and stored at -20°C.

Crude *E. coli* genomic DNA extraction

This protocol was used prior to diagnostic PCR of the allele present at a genomic locus. A single colony was extracted from an agar plate using a sterile wire loop, and suspended in 50µl of 10mM Tris-HCl in a 0.6ml centrifuge tube. It is essential not to

contaminate a PCR reaction with agar, so care was required removing *E. coli* from plates. The centrifuge tube was then incubated in boiling water for 4min. Cell debris was precipitated out by 2min fast spin on a benchtop centrifuge (9krpm), and 4µl of the resulting supernatant was used per 20µl PCR reaction volume.

Purification of PCR Products

DNA fragments ranging from 100 bp to 10 kb were purified from primers, dNTPs, salts and enzymes using the QIAquick PCR Purification Kit (Qiagen). This method of purification was used to ensure optimal conditions during an enzymatic reaction when, for example, incubation with a restriction endonuclease followed PCR. Like the Gel Extraction Kit the PCR Purification Kit uses silica-gel membranes. DNA binds to silica at pH $\leq$ 7.5 in the presence of a high concentration of chaotropic salts while other components of the PCR reaction mixture are found in the flow-through. The purification was carried out using the solutions provided by the QIAquick kit according to the protocol supplied by the manufacturer. 5 volumes of buffer PB were added to 1 volume of PCR reaction. After mixing thoroughly the solution was applied to a QIAquick column which was standing in a 2-ml collection tube. The column was centrifuged at 15 krpm for 1 minute at room temperature. The flow-through containing salts, dNTPs, primers and enzymes was discarded. The column was washed with 750 µl of buffer PE to remove residual salts and other PCR reaction components. Flow-through was discarded and the centrifugation step was repeated to remove all of buffer PE. The DNA was eluted from the column with 30 µl to 50 µl of 10 mM Tris-HCl and stored at -20°C.

Extraction of DNA Fragments from Agarose Gels

PCR products and DNA restriction fragments were extracted from agarose gels when it was necessary to select particular restriction fragments or PCR products for labelling of DNA probes, for ligation or DNA sequencing. DNA fragments were separated on horizontal agarose gels made from a Genetic Technology Grade (GTG) agarose (SeaKem Incorporated) which is special agarose for preparative electrophoresis of DNA. After gel electrophoresis was completed, the gel was placed on an UV transilluminator and the selected DNA band was excised from the agarose gel using a sterile scalpel. The gel slice was put into a sterile 1.5-ml Eppendorf tube

and was weighted. The QIAquick Gel Extraction kit (Qiagen) was used to extract the DNA from the agarose slice. First, 3 gel-volumes (w/v) of buffer QG were added to the gel slice and the gel/ buffer mix was incubated at 50°C until the gel had completely dissolved (approximately 5 min). The colour of buffer QG, which is yellow indicating a pH ≤ 7.5, should not change during this procedure. If it changed to orange or violet the pH was too high and had to be re-adjusted to pH 7.5 to ensure optimal binding of the DNA to the silica-gel membrane. If small DNA fragments (< 500 bp) were extracted, 1 gel-volume isopropanol was added to the solution to increase DNA recovery. Solutions containing longer DNA fragments were loaded onto the column without any modification. The columns were spun at 15krpm for 1 minute. The flow-through was discarded and 500 µl of buffer QG were added to the column to remove traces of agarose. The columns were centrifuged at 15krpm for 1 minute. The flow-through was again discarded and 750µl of ethanol-containing buffer PE were added to eliminate salts. The column was spun twice at 15krpm for 1 minute and the collection tube was emptied between centrifugations to ensure complete removal of buffer PE from the column. The column was placed in a sterile 1.5ml Eppendorf tube and 30µl of 10mM Tris-HCl buffer were added to the centre of the membrane to elute the DNA. After 1 minute the column was centrifuged for 1 minute at 15krpm. Since the lid had to be cut off the tube for centrifugation, the eluate was transferred to a new, sterile Eppendorf tube. The sample was stored at - 20°C.

DNA quantification

Measurement of the Concentration and purity of DNA

The calculation of the amount of DNA in a solution is based on the maximal UV-light absorption of DNA at a wavelength of 260nm. An absorption value of 1 measured at this wavelength equals 50 µg of double-stranded DNA per ml (Sambrook, Fritsch and Maniatis, 1989). The absorption spectrum between 200nm and 300nm was recorded as an indication of the purity of the DNA preparation. The quotient of the absorption values measured at 280nm and 260nm should be greater than 1.8 in a pure DNA solution.

Estimation of the relative proportion of two or more DNA bands in an agarose gel.

Vistra Green (Amersham pharmacia biotech) is normally frozen in the dark at -20°C in aliquots of DMSO stock solution. It was made up fresh each time by a 1:10,000 dilution in TAE for staining over a 40 minute incubation in the dark on a flat bed shaker. The gel was quickly surface rinsed before scanning in a Storm8600 under blue fluorescence mode at 900V and 100 pixel size (the minimum available). DNA quantification could then be undertaken from this image using ImageQuant software from Molecular Dynamics. Pharmacia claim bands of less than 20pg in agarose can be detected in this way.

Molecular Manipulation of DNA

DNA Restriction Digests

Unless stated otherwise DNA digestion was carried out in volumes of 20 µl using 2 U of restriction endonuclease per µg plasmid DNA with the incubation buffer provided by the manufacturer. The reaction mixture was incubated for 2 hours at the temperature recommended by the manufacturer (usually 37°C). If the DNA had to be modified subsequently (e.g. end-labelling, ligation, sequencing) the restriction endonuclease was heat inactivated at 65°C for 20 min.

The Polymerase Chain Reaction (PCR)

Primer design.

DNA sequence information was obtained from Entrez nucleotides (http://www3.ncbi.nlm.gov/Entrez/index.html). Primers were chosen using Gene Jockey™ and Primers 3™ software. All oligonucleotides were ordered from Oswel DNA service UK (www.oswel.com).

Usually, PCR was carried out using purified plasmid DNA or bacterial genomic DNA. PCR was carried out in a volume of 25 µl using 0.2ml PCR tubes in a PCR Express Hybaid Thermal Cycler. The reaction mixture consisted of:

2.5 µl 10 x Taq-Polymerase buffer (including 25 mM $MgCl_2$)

2.5 µl 2 mM dNTP-4 mix

5 µl DNA (plasmid DNA: 10-100 ng; genomic DNA: 200-400 ng)

0.5 µl Taq-Polymerase (2.5 Units)

1 µl primer 1 (20 pmole)

1 µl primer 2 (20 pmole)

13 µl sterile, distilled water

A master-mix containing all reaction components excepting template DNA was prepared on ice with Taq-DNA Polymerase being added last. Aliquots of 20 µl were dispensed into 0.2ml PCR tubes, which already contained 5 µl of DNA solution. In a control reaction the DNA solution was replaced by 5 µl of sterile distilled water. The solution was mixed well and centrifuged briefly at 9 krpm. All PCR reactions were performed as "hot starts" (the PCR block was pre-heated to 94°C before the samples were inserted). PCR programmes that were developed in the course of this project are shown in the table below. Optimisation of PCR programs was performed by comparing yields of identical reaction mixtures exposed to different annealing temperatures on a gradient thermal cycler (ABgene), as seen on an agarose gel. Subsequent optimisation of reaction buffers and adjuncts could then be performed if a higher yield was required. A typical program is shown below:

1 initial denaturation:		96°C for 2 min
30 cycles:		
	denaturation	94°C for 10 seconds
	annealing	55°C for 20 seconds
	elongation	72°C for 2.5 min
1 final elongation		72°C for 10 min

Radioactive End-Labelling of Double-Stranded DNA Fragments.

DNA fragments were labelled after incubation of plasmid DNA with restriction endonucleases (e.g. Eco RI), that produce 3' recessed ends which can be filled-in by Klenow enzyme. Because of the sequence of the overhang produced by EcoRI, [α-^{35}S]-ATP was used in the labelling reaction (Figure 2-1).

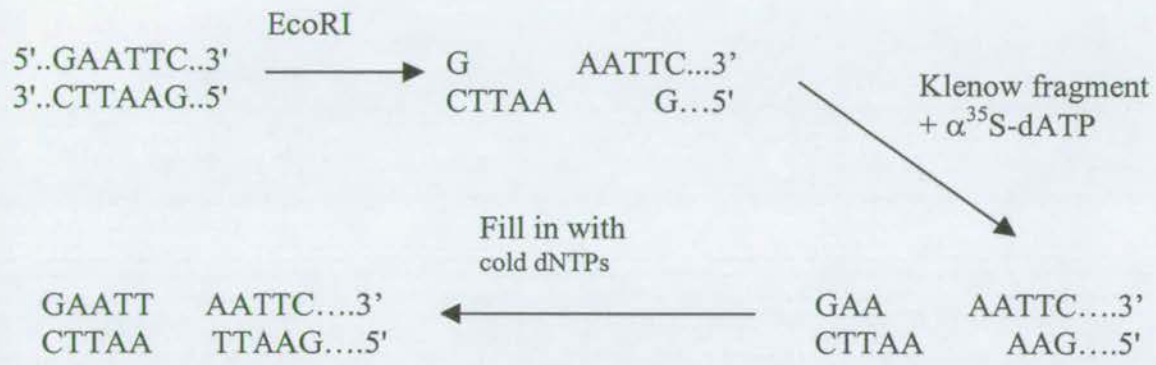


Figure 2-1. Radioactive labelling of EcoRI restriction fragments by incorporation of α^{35}S-dATP by Klenow enzyme.

During this project, radioactive end-labelling was used to determine the number of CTG trinucleotides in pUC18-derived plasmids which carry CTG repeat arrays of varying length. For this purpose, 17 µl of plasmid DNA prepared by the small scale method (see 2.2.4.2.1 Extraction of Plasmid DNA from *E. coli*) were mixed with 2 µl of 10 x EcoRI restriction buffer (New England Biolabs) and 1 µl of EcoRI (20Uµl^{-1}). After 2 hours of incubation at 37°C EcoRI was heat-inactivated at 65°C for 20 min. The samples were spun in a bench-top centrifuge at 9 krpm fpr 30 seconds. Without any further modification the samples were used in the following radioactive end-labelling reaction. Per sample, 1µl (10µCi) of [α-^{35}S] dATP and 0.5µl (2Uµl^{-1}) Klenow enzyme were combined in an Eppendorf tube. Of this labelling mix 1.5µl were added to 20µl of EcoRI-digested plasmid DNA. After incubation at room temperature for 5min, 2µl of the 2 mM dNTP-4 mix were added to the labelling reaction and incubation was continued for 10 min at room temperature. 3µl of polyacrylamide gel loading buffer were added to the reaction. Immediately, 20 to

25µl of this reaction mixture were loaded onto a native 5% Long Ranger gel to separate the labelled DNA fragments.

The predominant form of radiation from ^{35}S is β-particles (electrons), so work was carried out in controlled areas in a well-ventilated fume hood, behind a perspex screen. Two pairs of gloves, a lab coat, and personal dosemeter was worn at all times. Afterwards, the work area was monitored with a Mini Monitor Geiger counter. Samples with residual activity were disposed of in non-combustible radioactive waste bins.

Autoradiography

Dried polyacrylamide gels were placed in a Molecular Dynamics AutoRad cassette for overnight incubation, followed by scanning on a Molecular Dynamics Storm 8600 Phosphoimager. Quantification of relative DNA band strengths could then performed using ImageQuant software.

Separation of DNA species and fragments using gel electrophoresis

Agarose Gel Electrophoresis

The length of double-stranded DNA fragments was analysed by electrophoretic migration in horizontal agarose gels in 1 x TBE-buffer at aprozimately 5-8V/cm. Depending on the size of the DNA fragments, the concentration of routine electrophoresis grade agarose (Flowgen) was 1% to 2% (w/v) while higher percentage gels (3% to 4%) were made from Methaphor agarose (Flowgen). To pour an agarose gel, the appropriate amount of agarose was added to 1 x TBE or TAE buffer and mixed in a conical flask. TBE was found to be convenient for routine examination of PCR products, whilst scrutiny of plasmid multimerisation required TAE. The agarose was melted in a microwave oven at setting "High" for 1 - 3 min (depending on the gel volume), shaken vigorously, and inspected for undisolved agarose particles. When all agarose particles were melted, the flask was placed on a magnetic stirrer and left there to cool down until just cool enough to hold safely, then the gel was poured. The gel

was allowed to set for 1 hour at room temperature. Prior to loading, 1 x agarose gel loading buffer was added to each sample and mixed well. After electrophoresis, the DNA was stained using 2µl of 50 x ethidium bromide stock solution for a 100ml gel, diluted in just enough TAE/TBE buffer to cover the gel. 1 hour staining was followed by 30 min destaining in several changes of buffer. DNA was visualised using a C-62 BlackRay transilluminator (Ultraviolet Products Incorporated). Pictures of gels were taken using GRAB-IT™ software (Ultraviolet Products Incorporated).

Native Polyacrylamide Gel Electrophoresis

Native polyacrylamide gels were used to analyse length changes in the highly unstable CTG trinucleotide repeats. The gels were run in the Sequi-Gen® Nucleic Acid Sequencing Cell (Biorad) according to the manufacturer's instructions. The gel was prepared at least 3 hours before it was needed. For a gel measuring 40cm x 21cm x 0.4mm, 70ml of gel solution were required (20ml to seal the bottom of the glass plate sandwich and 50ml to pour the gel). In order to make a 5 % native gel containing 1.2 x TBE, 7ml of 50 % Long Ranger™ stock solution (Flowgen), 8.4ml of 10 x TBE and 54ml of sterile distilled water were mixed thoroughly but carefully in a glass beaker by slowly pipetting up and down. To seal the bottom of the glass plate sandwich 70µl of TEMED and 170µl of freshly prepared 10% (w/v) AMPS were added to 20ml of the gel solution. When the gel had set, 50µl of TEMED and 150µl of freshly prepared 10% (w/v) AMPS were added to the remaining 50ml of gel solution. This gel mixture was immediately poured between the two glass plates with the help of a 50ml plastic syringe. A square-toothed comb was introduced at the top of the gel. The gel remained at room temperature for at least 3 hours to set. After 3 hours the gel was removed from the casting tray and placed into the sequencing apparatus (Biorad). Upper and lower buffer tanks were filled with 700ml and 300ml of 0.6 x TBE buffer, respectively. Just before the samples were loaded the comb was removed and the slots were thoroughly washed with 0.6 x TBE using a syringe with a needle. Into each slot 20µl to 25µl of sample containing polyacrylamide gel loading buffer was loaded. The gel was run at a constant power of 35W at a temperature between 45-50°C until the bromophenol blue band of the loading buffer had reached the bottom of the gel. After the electrophoresis was completed, the gel was removed from the glass plate sandwich in VT fixing solution. It was then transferred onto wet blotting paper

(Ford Goldmedal), covered with Saranwrap and dried in a Biorad Gel Dryer (Model 583) at 80°C for 45 min.

Sterilisation of solutions

Autoclaving

Autoclaving was carried out by departmental services.

Sterile Filtering

For thermally unstable solutions, sterile filtering was carried out with sterile membrane filters (0.45μm) and sterile syringes in the sterile air behind a Bunsen's blue gas flame.

Chapter 3: Creation of strains containing complete knockouts of *recQ*, *recF*, and *yraN*.

Rationale

Mutations in several genes of interest were not available at the start of this study. A *recQ*⁻ strain was not present in the laboratory collection. In addition, the only *recF*⁻ allele contained an ampicillin resistance marker, making it unsuitable for transformation with pUC18-derived plasmids used as vectors for TR tracts. The development of a rapid means of engineering mutations into *E. coli* was considered a useful addition to the protocols practised within the research group, and could be used to disrupt the function of an unknown gene, *yraN*, which had sequence homology to resolvases, but had not been tested to give a phenotype.

Genes chosen for complete removal of activity

RecQ and its homologues

RecQ DNA helicase is known to be important in initiating homologous recombination and disrupting aberrant recombination intermediates. Its homologues are also thought to have functions in DNA postreplication gap repair pathways (Chakraverty and Hickson, 1999), re-initiation of replication forks, and cell cycle checkpoints. All of the *recQ* family genes encode 3'→5' DNA helicases of the DExH-box DNA/RNA helicase subfamily. There are at least five human homologues of RecQ helicase: RecQ1 and RecQ5 are low molecular weight helicases of approximately 600 amino acids with ubiquitous expression, whilst RecQ4, BLM, and WRN are large molecular weight proteins (approximately 1,400 amino acids). In *Saccharomyces cerevisiae*, the RecQ homologue is Sgs1, and in fission yeast it is rad12/rqh1.

Human *WRN* and *BLM*, and yeast *SGS1*, are hyper-recombinogenic when mutated, leading to increased chromosomal instability, and their domain structures are strikingly similar, suggesting similar functional roles in the cell (Ellis *et al.*, 1996). Human RecQ4 and BLM expression are upregulated at the G1/S phase transition, whereas RecQ1 and WRN expression are enhanced at the G2/M transition phase of the cell cycle (Kitao *et al.*, 1998). Yeast Sgs1p, and human BLM and WRN are all thought to have a similar role in a

replication checkpoint specific to S phase (Frei and Gasser, 2000). *S. cerevisiae* Sgs1, *S. pombe* Rqh1, and human BLM have been shown to participate in genome replication by disrupting Holliday junctions formed upon replication arrest (Doe *et al.*, 2000; Gangloff, 2000; Karow *et al.*, 2000).

The severe phenotype of many *recQ* homologue mutations, and the number of homologues in higher eukaryotes suggests an important and highly conserved function in cells. The substrate promiscuity of these helicases and the variety of roles they occupy in DNA replication and repair make them ideal as cell cycle checkpoint proteins for chromosomal integrity. Their evolutionary conservation, variety of DNA maintenance roles, and suggested interaction with DNA secondary structures makes them attractive candidates for possible involvement in TR dynamic mutations.

WRN

Werner syndrome (WS) is an autosomal recessive disorder characterised by genomic instability, the premature onset of age-related diseases, degeneration of the central nervous system, and high incidence of cancer, followed by death of the patient at an average age of 46 years. In common with the premature ageing phenotype, WS cells have a reduced replicative lifespan in culture. Genomic instability is observed in the form of chromosome breaks and translocations, creating multiple large deletions. The distance between sites of replication initiation is longer in WS cells (Takeuchi *et al.*, 1982).

WRN helicase (Werner Syndrome helicase) was shown to have an integral ATP stimulated 3'→5' exonuclease (Kamath-Loeb *et al.*, 1998). A recessed 3' strand is believed to be the substrate for digestion, and higher levels of hydrolysis are observed if the substrate contains a single 3'-terminal mismatch. WRN helicase has been shown to unwind RNA-DNA heteroduplex, but with lower efficiency than a DNA duplex (Moser *et al.*, 1999). A cDNA sub-species completely identical in sequence (but lacking a C-terminus nuclear localisation signal) was found to be expressed only in the testis (Wang *et al.*, 1998), suggesting a WRN helicase role in meiosis.

Recently, WRN has been demonstrated to unwind tetrahelical structures in the fragile X syndrome CGG repeat locus (Fry and Loeb, 1999). As WRN resolves DNA structures

including tetraplex DNA, triplex DNA, and Holliday junctions, one function of WRN may be to unwind secondary structures that impede cellular DNA transactions. Subsequent work showed that hairpin and G'2 bimolecular tetraplex structures of the fragile X expanded sequence, d(CGG)(n), effectively impeded synthesis by three eukaryotic replicative DNA polymerases (pol): pol alpha, pol delta, and pol epsilon. WRN facilitates pol delta (but not alpha or epsilon) to traverse these template secondary structures to synthesize full-length DNA products (Kamath-Loeb *et al.,* 2001). Alleviation of pausing by pol delta was also observed with *E. coli* RecQ, but not with UvrD helicase, suggesting a conserved function of RecQ family helicases in rescuing pol delta-mediated replication stalled by unusual DNA secondary structures.

BLM

The Bloom's syndrome gene product was first identified as a RecQ homologue in 1995 (Ellis *et al.,* 1995). Bloom's syndrome has an overlapping but distinct phenotype compared with Werner's syndrome. Immunodeficiency and short stature is combined with an extremely high incidence of cancer through chromosomal instability. Patients also have hypo- and hyper-pigmented, sun-sensitive skin.

Mutations in (BLM) result in extensive chromosome breakage and increased rates of sister chromatid interchange in somatic cells, but whereas WRN^- is thought to cause defective DNA replication initiation, BLM^- is thought to cause defective chain elongation. *BLM* maps to 15q26.1, and the elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15 (McDaniel *et al.,* 1992). The protein is a 158kD ATP and Mg^{2+}-dependent DNA helicase that unwinds in a 3'→5' direction (Karow *et al.,* 1997). There is some evidence that BLM protein detects specific types of DNA damage and regulates the cellular response via transcription of p53 and other proteins (Collister *et al.,* 1998). BLM has been shown to unwind quadruplex DNA as a preferred substrate (Sun *et al.,* 1998).

yeast Sgs1

Sgs1 helicase in yeast acts as a suppresser of illegitimate recombination. *Sgs1*$^-$ yeast display premature ageing, quicker sterility, shortened life span, and redistribution of the Sir3

silencing protein from telomeres to an enlarged and fragmented nucleolus. Strains lacking Sgs1 exhibit elevated levels of chromosomal mis-segregation during both mitotic and meotic divisions, possibly via an interaction with topoisomerase II (Watt *et al.*, 1995). Sgs1 helicase efficiently unwinds G-G mispaired DNA, such as that formed by telomere-telomere interactions that have the potential to lead to chromosome non-disjunction (Sun, Bennett, and Maizels 1999). Both Bloom's and Werner's syndrome genes suppress hyper-recombination in yeast *sgs1*$^-$ mutants (Yamagata *et al.*, 1998). Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases (Gangloff, Soustelle, and Faber 2000).

Fission yeast Rad12/Rqh1

Rad12 over-expression and under-expression mutants are sensitive to UV and gamma radiation, and the DNA synthesis inhibitor hydroxyurea. Chromosome segregation defects seen in *rad12*$^-$ mutants are associated with loss of S-phase and G-phase checkpoint control, which is thought to be the result of Rad12 coupling chromosome integrity to cell cycle progression (Stewart *et al.*, 1997). Rad12 was subsequently demonstrated to act upstream of Rad9 in checkpoint exit (Davey *et al.*, 1998). It has been suggested that Rad12 is involved in a recombination repair pathway which may involve replicative by-pass of UV-damaged DNA, necessary when the UV dimer cannot be removed (Murray *et al.*, 1997). Partial suppression of the fission yeast *rad12*$^-$ phenotype has been achieved by expression of a bacterial Holliday junction resolvase (Dabert and Smith, 1997). The authors of this report proposed that in the absence of Rad12, replication fork arrest results in the accumulation of Holliday junctions, impeding sister chromatid segregation.

RecF

In addition to the general importance of the RecF recombination pathway for ss gap repair (outlined in the introduction chapter 1) there are a number of other properties of RecF making it important in the study of the effect of recombination on TR stability in a plasmid system. Not least of which is the decreased plasmid recombination observed in *recF*⁻ mutants (Cohen and Laban, 1983; Kolodner *et al.*, 1985). Homologues in humans constituting an equivalent pathway to RecF for recombinase loading are Rad52, Rad51B/C/D, and XrCC2/3 (Cromie *et al.*, 2001).

There are also a number of studies linking RecF to replication in *E. coli*. Severely reduced levels of homologous recombination resulting from the introduction of a *priA*⁻ allele into a *recBC*⁻*sbcB*⁻ background indicate that RecF-mediated recombination can require replication (Kogoma *et al.*, 1996). RecF-dependent SOS induction pathways also require DNA replication (Sassanfar and Roberts, 1990), and *priA*⁺ is required for SOS induction in a *recF*⁻ strain. The combination of *priA*⁻ and *recF*⁻ null mutations is lethal in an otherwise wild type background (Sandler, 1996), suggesting overlapping functions of these two genes in replisome reassembly. This activity attributable to RecF is independent of RecO and RecR, as mutations in either of these genes do not reduce the viability of *priA*⁻ cells. The addition of a *dnaC809* allele restored cell viability but not SOS expression to *priA*⁻*recF*⁻ cells. *dnaC809* is a mutation known to suppress the UV sensitivity and recombination deficiencies exhibited by *priA*⁻, probably by being able to bind DNA substrates in the absence of PriA, allowing direct loading of DnaB (Sandler *et al.*, 1996).

It has been proposed that the UV hypersensitivity of *recF*⁻ *or recR*⁻ cells is due to an inability to re-initiate stalled replication forks, rather than a defect in recombination *per se*, as they have almost full proficiency for homologous recombination (Courcelle *et al.*, 1997). Indeed, *recF*⁻ or *recR*⁻ cells are far less UV sensitive in stationary phase when replication is less extensive. The replication-link of these two proteins is re-enforced by their genomic organisation. RecF is expressed from an operon containing the *dnaN* gene, and RecR is expressed from an operon containing the *dnaXZ* genes. Restart of DNA replication is delayed in *recF*⁻ cells, and ss gap filling is greatly reduced. It was subsequently demonstrated that cyclobutane pyrimidine dimers and 6-4 photoproducts are removed at full efficiency by the

UvrABC system in *recF*⁻ cells (Courcelle *et al.*, 1999). Resumption of replication from stalled replication forks after UV exposure requires both functional excision repair to remove the DNA lesions, and RecF to restart replication disrupted by these lesions. Unlike wild type and *uvr*⁻ cells, UV-irradiated *recF*⁻ cells display degradation of about half of their nascent DNA (Courcelle *et al.*, 1999). The specific function of the RecFOR complex at broken or stalled replication forks is believed to be to stabilise the fork structure when the replisome dissociates, protecting it from degradation by exonucleases and simultaneously promote re-initiation when possible.

"Thymineless death" is a form of rapid cell death occurring due to aberrant replication in the absence of thyimine. An elevated resistance to thymineless death is observed in *recF*⁻, *recO*⁻, *recJ*⁻ and *recQ*⁻ cells, whilst *recA*⁻ alleles have no effect (Nakayama *et al.*, 1988).

YraN

Blast searches (http://www.ncbi.nlm.nih.gov/blast) using the *yraN* sequence suggested a novel but conserved 14,8kDa protein with close homologues present within many eubacteria (figure 3.1). They exhibit a limited homology to Hjc including a nuclease superfamily metal binding signature, and catalytic lysine residue. *In silico* homology searches and structure prediction by other groups has similarly led to the conclusion that *yraN* is a member of a conserved family of "endonuclease fold" proteins possibly functioning as Holliday junction resolvases or DNA repair enzymes (Aravind, Makarova, and Koonin, 2000). YraN is predicted to consist of three alpha helices joined by two beta sheets, and appears as compact as a globular domain (results from "predictprotein@columbia.edu" and "Prodom.2001"). Homology searches place *yraN* in a superfamily of endonuclease fold enzymes containing the RecB family, the Mrr (a methylated-DNA specific restriction enzyme) family, the λ exonuclease family, and a family of nucleases characterised by the presence of a *Pyrococcus horikoshii*-type ATPase C-terminal domain. Specifically, *yraN* is a member of the Mrr family which also includes the archaeal Holliday junction resolvase from *Pyrococcus furiosus* (AHJR). All of the members of this family are thought to co-ordinate a divalent cation (Mg^{2+}) using a conserved metal-binding motif, and possess a catalytic lysine or arginine residue. The *yraN* promoter is a standard sigma70 binding site without a LexA box, suggesting that YraN is not SOS-inducible.

Figure 3.1: The UPF0102 family.

```
Consensus  1 AGEALAARWLESKGLRILERNWRS-RYGEIDIIARD--DRTIVFVEVKTRSGENYGGA-A 56
gi 6919986 7 AGESLVAAWLEQQGGKILQQRWRS-PWGEIDLITHFpdTKIIAFVEVKTRSGGNWDQGgL 65
gi 1176819 11SFEHQARLFLESKGLIFIAANQNF-KCGELDLIMND--KETIVFVEVRQRSHSAYGSA-I 66
gi 1176818 22AWEAQARRWLEGKGLRFIAANVNE-RGGEIDLIMRE--GRTTIFVEVRYRRSALYGGA-A 77
gi 6831723 11FGEAYAARWLATRGYIIITRNWRR-ATGEIDIIAQQ--DDTIVFVEVKTLRCTSYADL-A 66
gi 6919990 20YGETLAARRLTGAGMTVLERNWRCgRTGEIDIVARD-gDVLVVC-EVKTRRGGAFEHP-M 76
gi 1731288 15MGEALAVDYLTSMGLRILNRNWRC-RYGELDVIACDaaTRTVVFVEVKTRTGDGYGGL-A 72
gi 6919997 15MGEVFAVDNLTRMGLRGLHCNWRC-RYGECDVIASEtaHRTVVS-RLRSIAATVMEGS-R 71
gi 6919985  5EYEDLAARYLKSKGYQILGRNLRS-PYGEIDILAEF----KGRKVIVEVKGSETFFP--A 57
gi 2313953 14KAEEEACGFLKSLGFEMVERNFFS-QFGEIDIIALK----KGVLHFIEVKSGENFDP--I 66

consensus  57 EAVTPRKQKKLRRTARLWLASQD--EFDADCRIDVVIVR 93
gi 6919986 66 LAVNARKQEKIWQTANHFLASQPq-WSDWNCRFDVMIVF 103
gi 1176819 67 ESVDWRKQQKWLDAANLWLAKQNmsLEDANCRFDLIAFG 105
gi 1176818 78 ASVTRSKQHKLIQTARLWLARHNgsFDTVDCRFDVVAFt 116
gi 6831723 67 IIVGKRKQKRICETAKHFLASARe-YNHMCARFDVIVLR 104
gi 6919990 77 AAVTPDKAERLRRLAERWIQTHGg-APPGGVRIDLVGVL 114
gi 1731288 73 HAVTERKVRRLRRLAGLWLADQE--ERWAAVRIDVIGVR 109
gi 6919997 72 RSAPEQKVRWLRWLAGLWPANQD--EF---------- 96
gi 6919985 58 EKVTPHKLSKIIRTAYEVLGEEP-------KSIEVVVVY 89
gi 2313953 67 YAITPSKIKKMIKTIRCYLSQKD--P-NSDFCIDALIVK 102
```

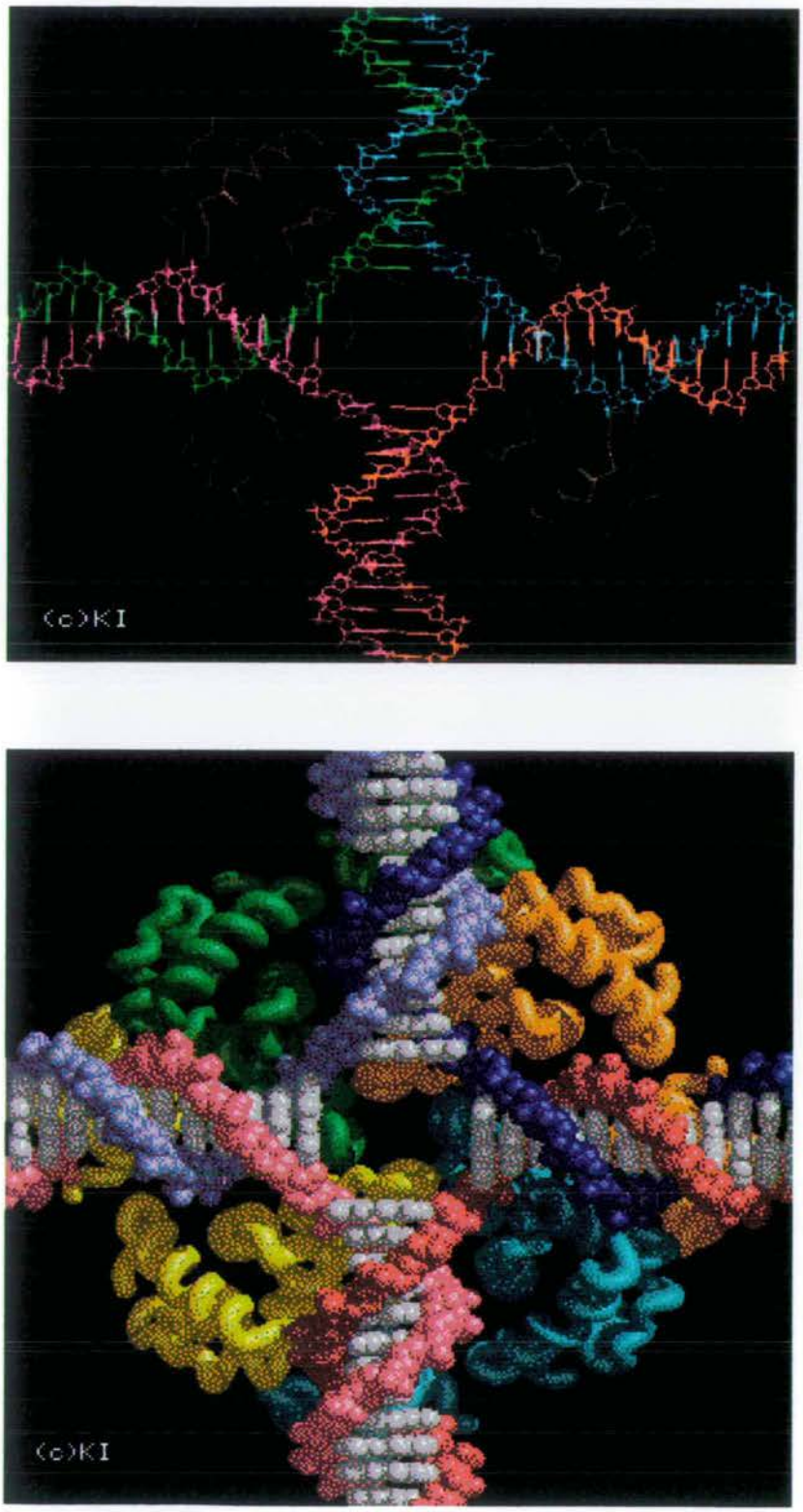
Synechocystis sp. gi 6919986
Haemophilus influenzae gi 6919986
Escherichia coli gi 1176819
Treponema pallidum gi 6831723
Streptomyces coelicolor gi 6919990
Mycobacterium tuberculosis gi 1731288
Mycobacterium leprae gi 6919997
Aquifex aeolicus gi 6919985
Helicobacter pylori 26695 gi 2313953

Holliday junction resolvases are notoriously difficult to identify as they are thought to have evolved function from at least four structurally distinct nuclease families as nature has taken advantage of various domains efficient in phosphodiester bond hydrolysis. RuvC for example, shares a fold and metal binding site with members of the RNase HI superfamily (Saito *et al.*, 1995). Many species encode a variety of HJ resolvases, some of which appear to have been acquired by horizontal gene transfer (Aravind, Makarova, and Koonin, 2000). Eukaryotic Holliday junction resolvases have proved difficult to identify, though their activities have been reported in yeast (Parsons and West, 1988) and mammalian cell extracts (Elborough *et al.*, 1990; Hyde *et al.*, 1994; Constantinou *et al.*, 2001).

Biochemical studies performed in the Malcolm White laboratory (University of St.Andrews) revealed that the protein is a homodimer in solution, suggesting that it binds a symmetrical substrate (figure 3.2). Unfortunately, determined attempts to crystallise YraN have proved unsuccessful thus far, so direct structural examination has not been possible. The diagrams shown below are a prediction of a model YraN structure binding a four-way DNA junction

(courtesy of Malcolm White), based on sequence alignment with other Holliday junction cleavage enzymes, and positioning of conserved catalytic residues.

Figure 3.2: Model of YraN bound to a 4-way DNA (Holliday) Junction.



Courtesy of Malcolm White. Unpublished.

Engineering of the *E. coli* chromosome using activated double strand break repair pathways

The rationale behind this protocol is to introduce a PCR-generated linear DNA duplex (with homology to a targeted region of the host chromosome), into a hyper-recombinogenic working strain. The introduced DNA undergoes ends-out recombination (two double strand break repair events) to become incorporated into the host genome, effectively using a double crossover to displace the wild type allele (figure 3.3). The concept of PCR-mediated disruption of genes was originally developed in yeast (Baudin *et al.*, 1993) because of the higher efficiencies of homologous recombination at short DNA homologies.

The DSBR pathways used here in *E. coli* require high levels of at least two activities: an exonuclease to generate a recombinogenic 3' overhang, and a recombinase to mediate homologous strand exchange between the invading and host duplexes. *recE* and *recT* are part of the Rac prophage in *E. coli* K12 strains, and provide a DSBR pathway (Kusano *et al.*, 1994) analogous to that provided by *red exo* and *beta*, which carry out general recombination events in bacteriophage lambda. RecE 5'→3' ATP-independent exonuclease (ExoVIII) resects duplex DNA to give free 3' ends (Muyrers *et al.*, 2000). This activity normally provides a very similar functional role to RecBCD in conjugational recombination (Clark *et al.*, 1994). *recT* genetically partly overlaps the c-terminal of *recE*, and its transcription product is known to form spiral filaments when bound to single stranded DNA, and to mediate pairing of homologous DNA sequences (Kolodner *et al.*, 1994) rather like RecA*. RecT bound to ssDNA is able to form D-loops by strand invasion (Noirot and Kolodner, 1998), allowing RecA-independent recombinational initiation of replicative DSBR.

To confirm the principles of this protocol, the pioneering experiments studied homologous recombination between linear and circular DNA (Zhang *et al.*, 1998; Muyrers *et al.*, 1999), and were later expanded to experiments targeting chromosomal locations (Yu *et al.*, 2000). The linear DNA produced by PCR carried a selectable marker gene flanked by sequences with homology to the plasmid DNA, thus dictating the specificity of the integration site into the plasmid. Initial experiments used sixtymer primers consisting of forty two nucleotides of 5' homology, and an eighteen base sequence to prime the amplification of the

chloramphenicol resistance gene. In subsequent experiments a direct relationship was found between homology arm length and recombination efficiency, with 27bp considered to be an absolute minimum. The recombination events were always found to be precise, but demanded high concentrations of DNA construct entering electrocompetent cells. The necessity for RecT and RecE in these recombination events was proved using an inducible expression plasmid to correlate the degree of RecET expression to the frequency of the desired recombination event. No targeted mutations were produced in the absence of RecT and RecE.

The activated *recET* pathway was originally observed in a *recBCsbcA* strain. Such a strain is also convenient for transforming purposes, since the RecBC enzyme is a highly processive exonuclease which would degrade incoming DNA, and hence reduce transformation efficiencies. *sbcA* mutations are thought to act as promoters for *recET* expression since they are dominant and cis-acting (Fouts *et al.,* 1983). Recombination with introduced linear fragments has also been reported in *recBC sbcB*$^-$ hosts (Winans *et al.,* 1985) and in wild type cells by the use of intentionally positioned *chi* sites (Dabert and Smith, 1997). However these often require complex artificial recombination substrates and long regions of homology. ET cloning is possible in *recBC*$^+$ strains using a plasmid over-expressing RecE, RecT, and Gam (an inhibitor of RecBC).

This system does not require the traditional restriction enzyme and ligation procedure for genetic engineering, so enables chromosome modification independent of the location of restriction endonuclease sites, eliminating any need for their creation by site directed mutagenesis. It therefore offers a very rapid two-step method for genetic modification. However, it is a relatively new protocol for engineering the chromosome of *E. coli* and had not previously been attempted in this laboratory.

Two working strains were chosen for initial attempts at a PCR-based *recQ* knockout. The recombination mechanism leading to gene conversion is believed to be very similar for the two pathways employed.

1. The first was JC8679 (DL49; AB1157 with $recB_{21}$ $recC_{22}$ $sbcA_{23}$). AB1157 was used as a negative control strain as it does not have an activated ET pathway for recombination.

2. The second strain DY329 (DL1218) expressed the analogous Lambda Red genes (*Exo* and *Beta*) together with *gam* (a repressor of RecBCD / ExonucleaseV) in a λp_L operon under the control of a temperature-sensitive λ *cI*-repressor (allele *cI857*). A deletion extending from *cro*, through the *attR* site, and into the *bioA* gene of the host removes the lytic genes of the prophage. The absence of *cro* allows the p_L operon to be fully de-repressed when the cI-repressor is inactivated at 42°C.

In both pathways the experimental rationale is the same. In the flow chart below, *recQ* is used as an example target sequence to be replaced by a kanamycin resistance marker.

Figure 3.3: The overall strategy of recombination-mediated chromosome engineering.

Figure 3.3a: Preparation of the transforming cassette by PCR amplification of the kanamycin resistance gene (from pUC4K) using primers flanked by sequences identical to those upstream and downstream of recQ in the *E. coli* chromosome. The same transforming cassette preparation could be used for both hyper-rec strains.

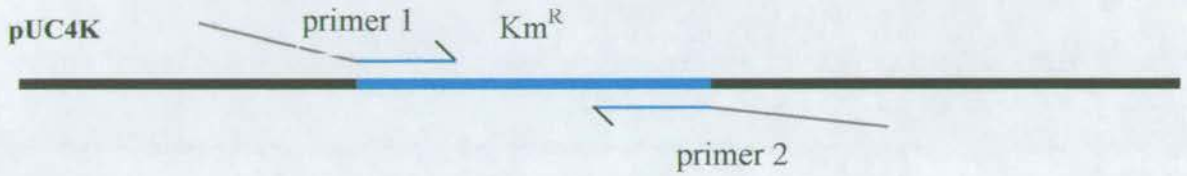


Figure 3.3b: The resulting cassette was introduced into an appropriate working strain by electroporation.

Figure 3.3c: An "ends-out" recombination reaction occurs. The working strains possessed either inactivated (*recB*⁻) or suppressed (Gam) ExonucleaseV activity. The activated *recET* or Lambda *ExoBeta* pathways then mediate homologous pairing and strand exchange between the two flanking regions of the construct and the chromosome.

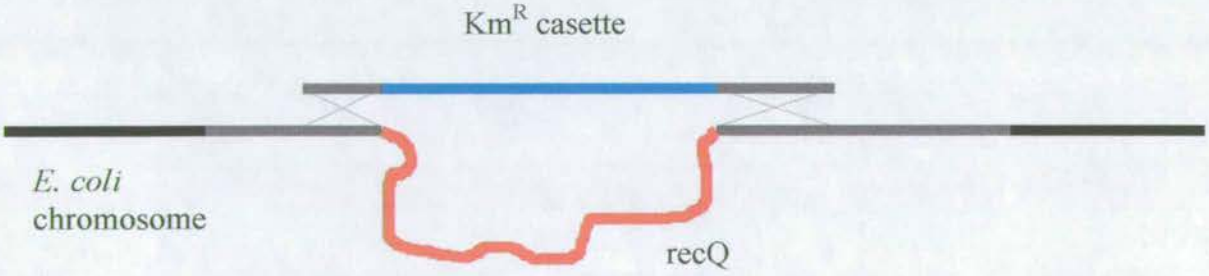


Figure 3.3c: The kanamycin resistance gene has replaced *recQ* on the chromosome.

Production of transforming DNA constructs using PCR

Plasmid pUC4K was extracted by Qiagen miniprep kit from host strain DL1083. A double digest was performed using *Eco*RI and *Bam*HI. PCR was performed in 50µl reaction volumes on a rapid thermal cycler (Idaho technology) using opened pUC4K as template DNA. PCR was optimised using a Stratagene Opti-prime buffers set and a Hybaid gradient thermal cycler.

Typical conditions:	reagents (all Boehringer Manheim/Rosche)
96°C for 2 min	5µl dNTP
94°C for 10 seconds }	0.5 µl Taq polymerase
55°C for 20 seconds }×30 cycles	5µl ×10 taq pol reaction buffer
72°C for 2.5 min }	1.3µl each primer
72°C for 10 min	34.9µl sterile milliQ water.

PCR product band size and yield were confirmed by loading 8µl samples onto 1% agarose gels. PCR products were pooled and purified using Quiaquick PCR purification kits. *Dpn*1 digest (at 36°C for 1hour, NEB buffer 4) was performed to digest pUC4K template DNA but retain the unmethylated DNA construct. This is an essential step, since any re-circularised plasmids present in the transforming mixture would enter and be maintained in cells at a far greater efficiency compared to short linear dsDNA, which would necessitate the screening of many kanamycin resistant colonies to find a successful recombinant. *Dpn*1 and any residual protein or salts were removed by phenol:chloroform extraction, followed by ethanol precipitation and vacuum drying. The DNA construct was then re-suspended in TE buffer at a concentration of $0.5\mu g\mu l^{-1}$. Once the DNA was pure and concentrated, the electrocompetent cells could be prepared, and electroporation performed on the same day (it was found that electrocompetent cells have a lower transformation frequency after freezing than when fresh).

Preparation of JC8679 (DL49) and activated DY329 (DL1218) electrocompetent cells.

DL1218 was grown at 32°C. It contains a defective Lambda prophage containing the *exo*, *bet* and *gam* genes under the control of a temperature-sensitive cI-repressor. Thus activation of the Red Gam recombination pathway would occur transiently after 10 min at 42°C. DL1218 which was not heat activated was used as a negative control. No activation was required for RecET expression in DL49, so DL68 acted as an isogenic negative control. Any kanamycin resistant colonies from either DL68 or unactivated DL1218 would indicate contamination of the transforming mixture by pUC4K.

Electroporation was performed as described in materials and methods (chapter 2) using 2µl of transforming DNA cassette suspended in TE, and 40µl of electrocompetent cells. The competence of the electrocompetent cells was tested using pUC4K. The worst transformation frequency found during this set of experiments was four times as great as the recommended minimum level of competence for this protocol (http://www.embl.heidelberg.de/ExternalInfo /stewart/Etcloning-textonly.html).

Selection of mutant colonies

Electroporated cells were plated out on L-agar + kanamycin ($50\mu gml^{-1}$) and incubated overnight. Colonies were individually toothpicked onto kanamycin and kanamycin with ampicillin grid plates. Since pUC4K encodes ampicillin resistance as well as kanamycin resistance, colonies which grew on both of these antibiotics were rejected. Colonies which grew well on kanamycin but which were susceptible to ampicillin were selected for further screening by diagnostic PCR through the *recQ* locus.

Individual colonies were boiled for 4 min in 10mM Tris:HCl, centrifuged, and diluted ten-fold. 2µl of this solution was then used as template DNA for PCR with primers binding outside the flanking sequences used for homologous pairing. The PCR product given by $recQ^{+}$ was 2039bp, but the product given by $recQ::Km^{R}$ was 1396bp. The PCR product lengths expected (and obtained) for $recF^{+}$ was 1145bp, and for $recF::Cm^{R}$ was 1010bp. The

values for *yraN*$^{+}$ was 1043bp, and 1964bp for *yraN*::*Km*^{R}. Phenotypically, recombination deficiencies in the *recF*$^{-}$ and *recQ*$^{-}$ strains were confirmed by pUC18 plasmid dimerisation frequencies similar to those of *recR*$^{-}$ and *recJ*$^{-}$ respectively. *YraN*$^{-}$ cells were subsequently tested for resolvase activity, mainly through the use of UV exposure.

Results and discussion of PCR cloning experiments

Despite repeated attempts at a *recQ* knockout, DL49 did not yield the desired recombinant. In contrast, strain DL1218 gave a successful *recQ* knockout on the first attempt, so was used alone in the subsequent knockouts of *yraN* and *recF*. Throughout these experiments the Red Gam system in DL1218 was found to be an effective and rapid method of producing clearly defined mutations in specific genes. Any reservations regarding the introduction of a *recQ*$^-$ mutation into DL49 proved to be unfounded, as subsequent P1 transduction from DL1285 was successful.

Unlike strains carrying a point mutation, the strains resulting from this protocol would be completely devoid of any of the activities of the targetted gene. Since none of the wild type protein could be produced, no possibility of partial catalytic activity, DNA substrate targeting, or protein complex formation need be considered when interpreting the results of subsequent experiments. The mutation was also permanent within strains carrying single-copies of the gene, unlike transposable element insertion mutants, which may revert back to partial or wild type activities upon excision of the transposon. Transposon excision and point mutations linked to marker genes have also lead to false positives when moving a mutant allele from one strain to another by P1transduction, as the antibiotic resistance marker may be transferred, but not the mutant allele itself. No such problems were observed when moving the new alleles created in this work. Also, any selectable marker could be used in the original DNA construct to replace the gene of interest, a very useful characteristic considering the subsequent production of a double or triple mutant strain by P1 transduction.

It is interesting that the Red pathway functional in DL1218 gave a *recQ*$^-$*Km*^{R} mutant at the first attempt, whilst the RecET pathway functional in DL49 refused to yield the desired mutant, even after many attempts. One striking difference between the two strains is the transient versus permanent inactivation of RecBCD. DL49 carries *recB*$_{21}$ *recC*$_{22}$ mutations, whilst DL1218 achieves a temporary inactivation of RecBCD through expression of Gam. It is certainly thought to be the case that up to 70% of cells in a recB$^-$ strain are non-viable; however, this reduced viability is recovered in *recB*$^-$*sbcA*$^-$. DL49 cell viability was normal. It is perhaps possible that the permanent inactivation of RecBCD, or permanent activation of the RecET pathway caused viability problems when combined with the desired *recQ*$^-$ allele, but this combination of mutations is certainly not lethal, since *recB*$^-$*sbcA*$^-$*recQ*$^-$ has been

observed previously (Kusano *et al.*, 1994). A *recB21 recC22 sbcA23 recQ::Km*^{R} strain was created in this work by P1 transduction of the *recQ::Km*^{R} allele into DL49, but at reduced frequency compared to transduction into a wild type strain. Recent work by Poteete and Fenton (2000) has indicated that removal of *recQ* causes a twenty-fold decrease in recombination mediated by the λRed genes. However, the λRed recombination pathway is transiently operative in DL1218 cells, unlike the constitutive RecET pathway in DL49. Where as DL1218 would be fully proficient in normal recombination after Gam expression is curtailed, DL49 must henceforth rely on the hindered RecET pathway available to it.

It is also likely that the required recombination event simply occurs at a greater frequency in DL1218 than DL49, perhaps as a result of the elevated expression of the recombination proteins, or due to the greater effectiveness of the λRed proteins in catalysing the reaction. Interestingly, Yu *et al.*, (2000), also considered recombination in *recBC sbcA* (Zhang *et al.*, 1998) to be inefficient between regions of relatively short homology (30-50bp). In contrast, it is claimed that under optimal conditions the Red pathway allows an efficiency of recombination approaching 0.1% of cells normally surviving a standard electroporation. Other reports claim that the engineered Red pathway displays 10-100 fold the frequency of transformation with linear DNA substrates relative to *recBCD sbcBC* and *recD* strains (Murphy 1998). Other versions of the engineered λRed pathway, such as *recBCDΔ::red* and *galKΔ::λ red gam* have also been found to be hyper-rec (Murphy *et al.*, 2000). Recently, a more sophisticated system has been developed which allows Red-mediated recombination to replace a chromosomal gene with an antibiotic resistance gene flanked by an FRT repeat (Datsenko and Wanner, 2000). Expression of FLP is then used to remove the resistance marker, leaving a small FRT scar. The Red and FLP activities are transcribed from temperature-sensitive replicons on low copy number curable helper plasmids. This plasmid-based system may be suitable for transfer into other species of bacteria.

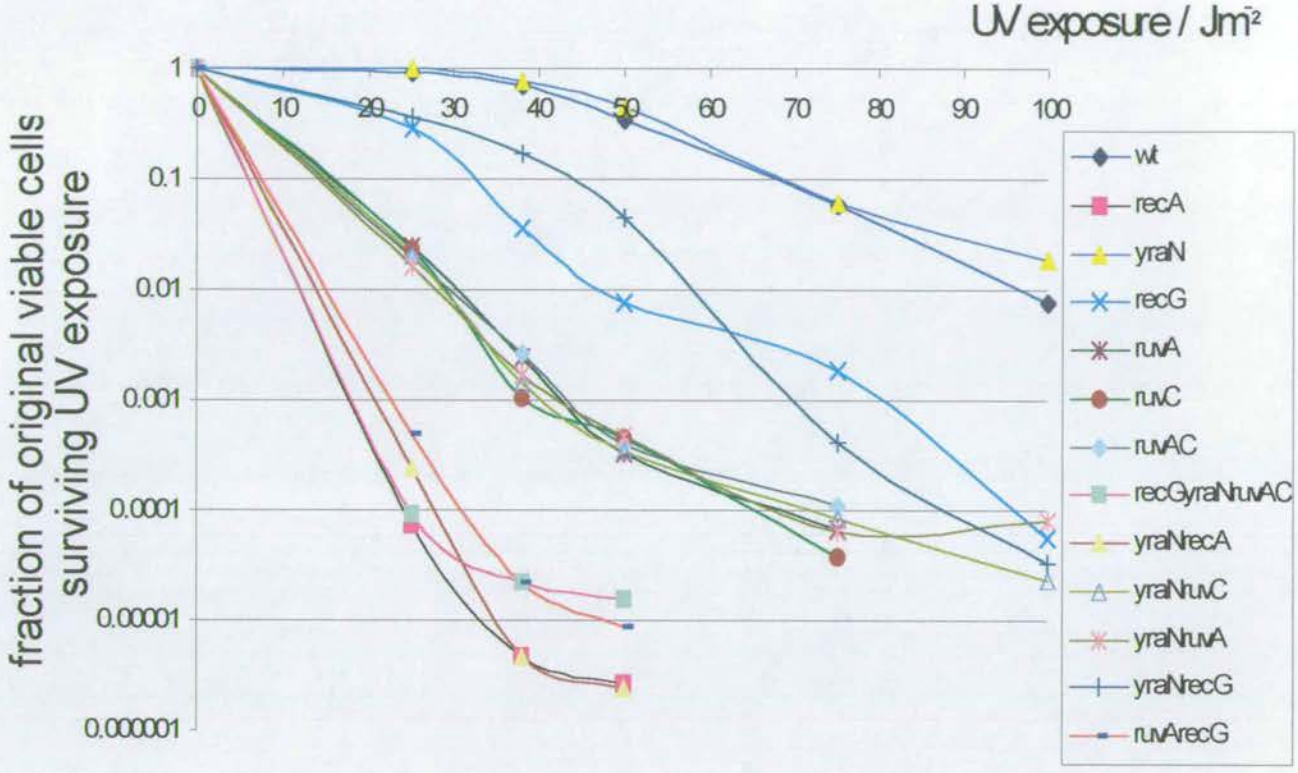
YraN is not a novel resolvase

To test whether YraN functions in Holliday junction resolution, phenotype testing was carried out not only in wild type cells, but also in cells deficient in the two known pathways of Holliday junction resolution in *E. coli*, namely the RuvABC and RecG pathways. Therefore the *yraN*⁻ mutant was examined by UV survival studies as per Lloyd 1991, in wild type, *ruvC*⁻, *recG*⁻, and *ruvC*⁻*recG*⁻ double mutants. 400µl of cells in log phase growth were harvested from antibiotic supplemented LB and re-suspended in phosphate buffer (which absorbs far less UV light). 7µl aliquots of serial ten-fold dilutions of these cell solutions were spotted onto LB agar and allowed to dry. This was performed on a perfectly flat table, and the spots were distributed as far apart as possible (maximum 16 spots per plate) to prevent spots merging. Several identical plates were prepared, and each was exposed to UV for different lengths of time using an UV Stratalinker. Colony counts were performed after overnight incubation. Only dilutions giving rise to less than a hundred colonies were counted, equivalent to two spots per strain for each timed UV exposure. The results were found to be extremely reproducible.

The results shown in graph 3.1 (below) demonstrate that the *yraN*⁻ allele had no effect on UV survival in *ruv*⁻ or *recG*⁻ single mutants, and the fully HJ resolvase deficient strain $\Delta recG_{263}$::*Cat* $ruvAC_{65}$ double mutant. This suggests that YraN does not have any HJ resolvase activity. The identical survival of *yraN*⁻ single mutant with wild type cells implies that YraN does not have a dominant role in cell survival or growth after UV exposure. This is reinforced by the identical UV survival characteristics of *recA*::Cm^R cells and *recA*::Cm^R *yraN*⁻Km^R, suggesting that YraN does not have the ability to aid repair of UV-damaged DNA by another pathway independent of homologous recombination.

Similar results were obtained in X-ray survival experiments carried out in an identical manner to the UV results shown here. X-rays also induce DSBs in DNA. Further examination using other mutagens was not pursued.

Figure 3.4: A mutation in yraN does not modify the UV survival of resolvase-defficient E. coli



Subsequent work in Malcolm White's laboratory has shown that purified YraN exists as a dimer in solution, suggesting binding preferences for a symmetrical substrate. However, gel-shift assays could not detect binding to four-way DNA junctions, bulges, duplex DNA, replication fork structures, or bubbles. This *in vitro* work alone does not rule out the possibility that YraN requires accessory partner proteins for correct function.

The broad distribution of *yraN* homologues in bacteria suggests that it is not a restriction enzyme involved in degradation of phage DNA. The high degree of conservation among its homologues suggests an important role in the cell. It may be a redundant nuclease, possibly operating in a pathway of DNA repair other than recombination. In an attempt to determine any DNA-repair phenotype for *yraN* cells, DL1333 was included in subsequent multimerisation and end-labelling studies performed in this thesis, but generated similar results to those of the wild type. Growth rates and transformation with pUC18-derived plasmids (including those with TR tracts) were indistinguishable from wild type *E. coli*. P1 phage growth on *yraN* wild type AB1157 was normal, and P1 transduction frequencies when

moving the *yraN*::Km^R allele into other backgrounds were perfectly normal when creating the following double and triple mutants:

yraN::Km^{R-} w.t AB1157
$recB_{21}$ $recC_{22}$ $sbcA_{23}$ *yraN*::Km^R
$recB_{21}$ $recC_{22}$ $sbcB_{15}$ *yraN*::Km^R
$recF_{143}$ *yraN*::Km^R
recF::Cm^R *yraN*::Km^R
$ruvA_{60}$::*Tn10* *yraN*::Km^R
$ruvAC_{65}$ *yraN*::Km^R
Δ*ruvC*::Tc^R Δ*xerC*::Cm^R *yraN*::Km^R
Δ$recG_{263}$::*Cat* *yraN*::Km^R
recA::Cm^R *yraN*::Km^R
Δ$recG_{263}$::*Cat* $ruvAC_{65}$ *yraN*::Km^R
mutS::Tc^R *yraN*::Km^R

The lack of additional UV sensitivity of *yraN*::Ω in wild type, *recA*::Ω, or *recG ruvAC* backgrounds as well as normal growth and recovery of P1 transductants strongly suggests that YraN does not have an important role in recombinational repair.

Summary

Three target genes (*yraN, recQ,* and *recF*) were chosen for knockout production by a new method making use of activated double strand break repair pathways. The practical success of the Lamda Red /Gam pathway compared to the RecET pathway is probably due to transient but powerful expression of the key recombination proteins and RecBCD inhibitor, resulting in exaggerated recombination frequencies coupled with good cell viability. Mutant production in DL1218 was found to be rapid and accurate. Very concentrated solutions of purified transforming cassette gave improved frequencies of mutant generation, but the key factor in mutant production was found to be the preparation of good electrocompetent cells.

Mutant production was verified using antibiotic selection, modified diagnostic PCR product size, and decreased plasmid multimerisation for *recF::Cm*^{R} and *recQ::Km*^{R}. The inability of mutants to revert back to functionality, and the complete absence of transcript activities, gave added confidence in subsequent experiments using the mutant alleles. The *yraN::Km*^{R} allele was checked for UV sensitivity in a variety of resolvase and recombinase deficient backgrounds. Contrary to encouraging protein structure predictions, these experiments did not produce evidence indicating a direct involvement of YraN in recombinational repair.

Chapter 4: Multimerisation of plasmids containing mouse $(CAG/CTG)_{43}$ trinucleotide repeat tracts

Aims

The formation of plasmid dimers and higher multimers occurs predominantly by recombination. In the series of experiments contained within this chapter, plasmid multimerisation in *E. coli* is used as a physical assay for the ability of trinucleotide repeats to induce recombination compared to pUC18 control. Since different recombination substrates have different preferred outcomes of Holliday junction resolution (figures 4.1 and 4.2), different recombination substrates would be expected to have predominantly crossover (leading to dimeric plasmids) or non-crossover (leading to monomeric plasmids) products. A variety of recombination deficient strains were tested to assess whether TR tracts could stimulate plasmid dimerisation in the absence of normal recombination pathways. It was expected that in certain *rec*$^-$ mutants differences in plasmid dimerisation rates attributable to TR tracts would enable extrapolation to a limited number of original presynaptic substrates. In this way the mechanisms of recombination at TR tracts was examined.

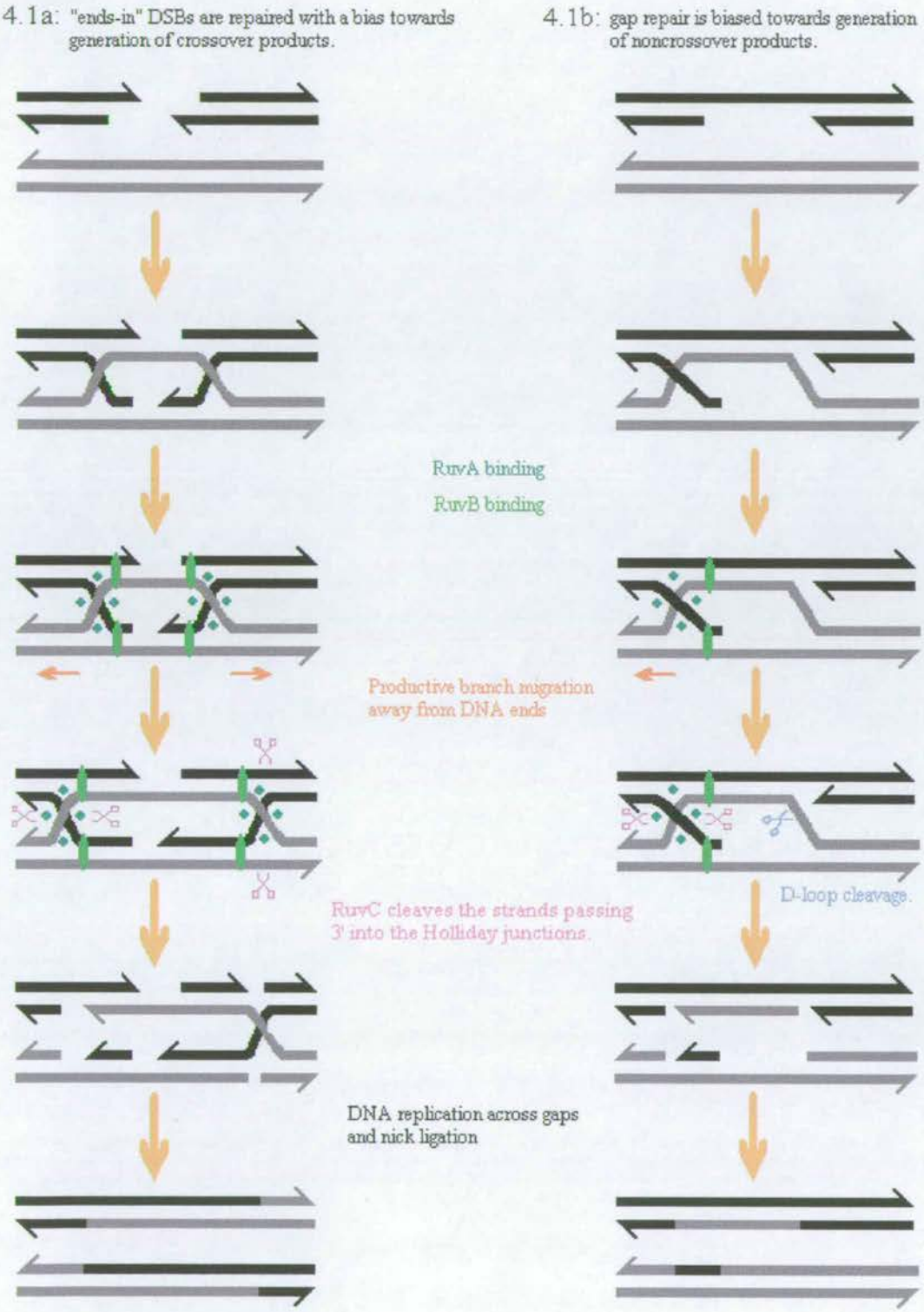
Introduction to plasmid multimerisation

Plasmid recombination occurs by the RecF pathway of presynapsis. Plasmid recombination mediated by RecBCD and independent of the RecF pathway has only been observed in plasmids bearing a *chi* site (Zaman and Boles, 1996). In recombination generating a crossover product, the two recombining duplexes are physically spliced together to create novel arrangements of flanking genetic markers on the same strand, whereas noncrossovers retain the unmodified (parental) configuration. The physical crossover of a strand from one duplex to another means

that only crossover products of recombination lead to dimeric plasmids. It was previously estimated that resolution of Holliday junctions was random. This would give 50% crossover and 50% non-crossover products, depending on which two alternative strands are cleaved. However, recent work suggests that resolution of the majority of Holliday junctions is deliberately biased towards strands which would not give crossovers and subsequent dimeric products (Cromie *et al.*, 1999). Van Gool and co-workers (1999) demonstrated that the orientation in which the RuvABC resolvasome was loaded onto the DNA junction resulted in up to a 50-fold bias in the resolution of the Holliday junction. The orientation of RuvAB loading also determines the direction of branch migration. Branch migration occurs as the RuvB rings pull DNA through themselves. For branch migration to productively increase the length of heteroduplex, the RuvAB complex must push the Holliday junction away from the site of initiation, which means that the RuvAB complex must have loaded onto the DNA in a specific orientation. RuvC cleaves the strands passing 3' through RuvB into the Holliday junction, so the generation of a crossover or non-crossover product is determined by the initial loading of RuvAB onto the postsynaptic substrate. Please note that if this initial RuvAB loading occurs in an orientation that pushes the Holliday junction towards the site of initiation, then branch migration would eliminate the heteroduplex, and strand invasion would have to be re-initiated for recombination to proceed. There may be a small minority of recombinational events (10% according to Cromie, 1999) that result from "non-productive" placement of the RuvABC resolvasome, with RuvC cleavage of the DNA strands occurring before the region of heteroduplex is completely removed.

By following the rules of RuvAB placement to generate productive branch migration, and RuvC cleavage of specific DNA strands, various recombination substrates can be predicted to have preferred crossover or non-crossover products. For standard recombination events, an ends-in DSBR would result in crossover products, but ss gap repair would produce non-crossover products.

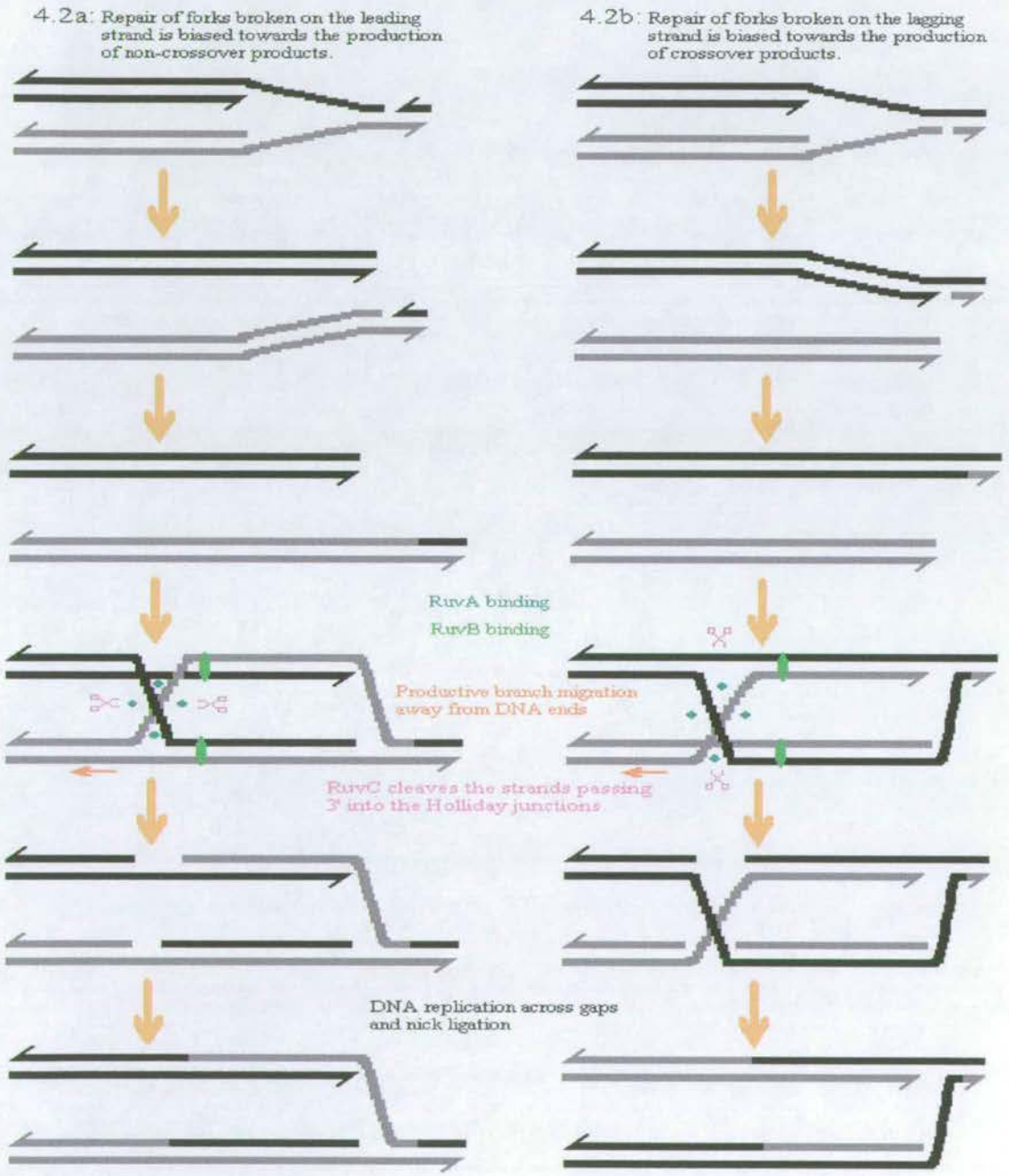
Figure 4.1: Processing of specific recombination substrates is biased towards crossover or non-crossover products.



adapted from Cromie and Leach, 2000.

Recombination initiated at replication forks by a leading strand break or paired nascent strands after fork regression, would result in non-crossover products, but recombinational repair of a break in the lagging strand would result in crossovers (figure 4.2).

Figure 4.2: Crossover and non-crossover products resulting from recombination at broken replication forks.



adapted from Cromie and Leach, 2000.

Multimers of natural multicopy plasmids, such as ColE1, are resolved to monomers by the Xer-*cer* site-specific recombination system (Summers, 1998). Because pUC18 does not contain a *cer* site, only standard recombination could occasionally convert dimers back to monomers. The equilibrium concentration of dimers is proportional to the production of dimers and is essentially independent of the low rate at which homologous recombination converts dimers to monomers (Summers *et al.*, 1993). Linear dimers possess the intrinsic ability to out-replicate monomers, as they posses two origins of replication which independently initiate forks capable of replicating the entire dimer. This adds to the sensitivity of this assay, as linear dimeric products of recombination will account for larger proportions of the total plasmid population in subsequent bacterial generations.

Recombinational models of plasmid dimerisation

The genetic requirement for the RecF presynapsis pathway in plasmid recombination (Kolodner, Fishel and Howard, 1985) has made ss gap repair an attractive model for plasmid dimerisation. It should be noted however that only a small minority of recombination events mediated by ss gap repair would be predicted to result in crossover products (Cromie and Leach, 2000). Thus a large amount of background recombination in the vectors used in this study might be occurring independent of TR tracts and would not be detected by this assay. On the other hand, recombination initiated at DSBs on the lagging strand of plasmid replication could result in a majority of crossover products. If a gap is present within a template strand (such as after MMR), a double strand break would be produced as a replication fork passed through. DSBs may also be formed by SbcCD processing of hairpin structures in regions of ssDNA.

In vitro work has shown that TR tracts inhibit DNA synthesis along a ss DNA template (Kang *et al.,* 1995; Usdin and Woodford, 1995; Kamath-Loeb *et al.*,2001), and TR tracts have demonstrated an ability to stall replication forks *in vitro* (Samadashwily, Raca, and Mirkin, 1997). Work by Michel would suggest the formation of DS ends (DSEs) at regressed replication forks. However recombination initiated by these DSEs would produce few crossover products so it is unlikely that

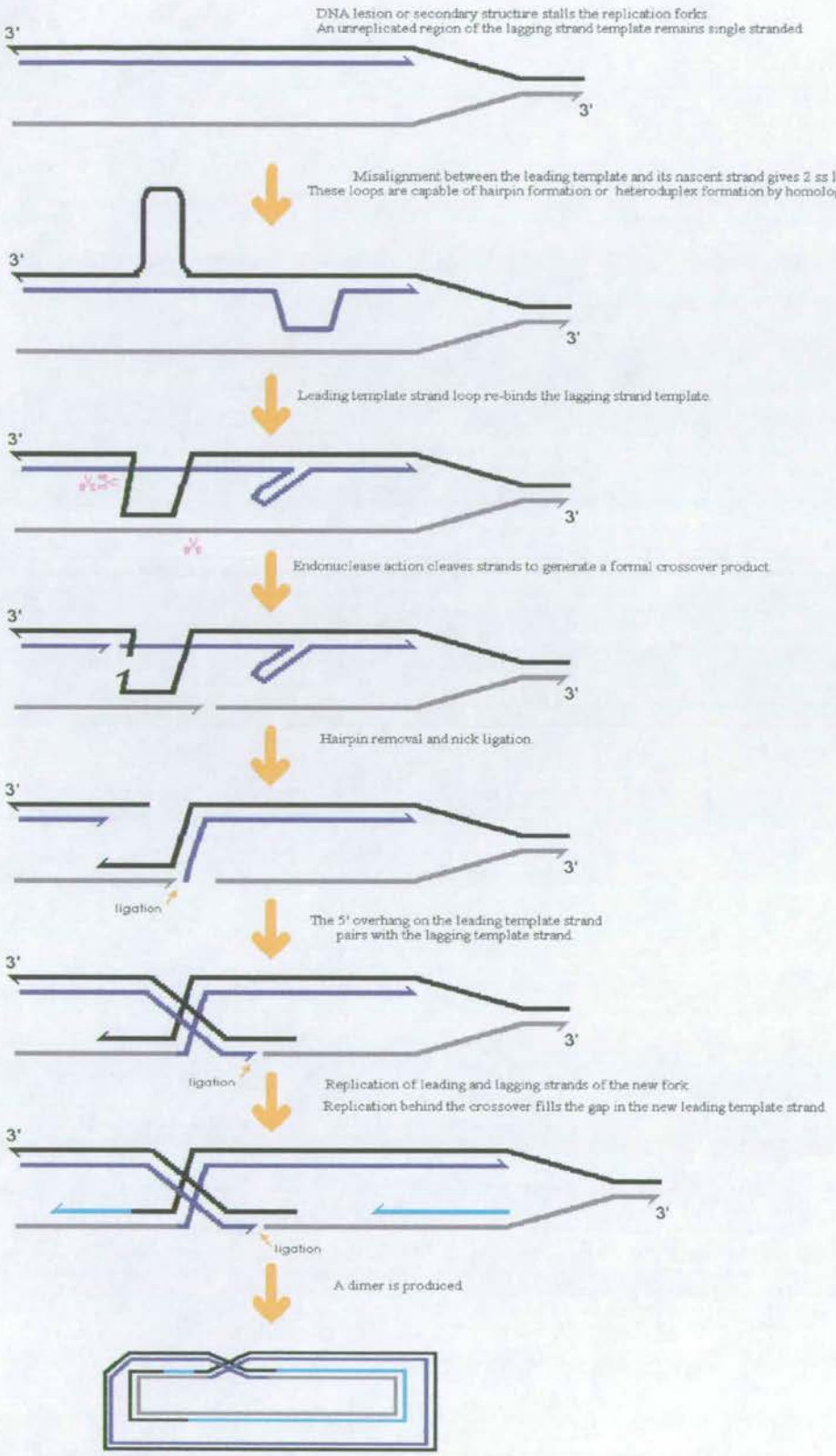
plasmid dimerisation would result from recombination initiated at replication forks stalled at TR tracts.

The high copy-number of pUC18-based plasmids also allows for dimer production by inter-plasmid recombination by conventional pathways. A double strand break upstream of a replication fork might be repaired by homologous pairing with an intact plasmid. However, it is expected that the absence of *chi* on a plasmid suffering a DSB would lead to complete degradation of the plasmid by the highly processive RecBCD enzyme. An exception to this may be in circumstances when the DSB is not a suitable substrate for RecBCD, such as if a long ss overhang is coated in SSB.

RecA-independent mispairing between direct repeats

Plasmid dimerisation could occur as a result of RecA-independent sister strand exchange events at stalled forks. A leading strands misalignment model proposed by Bi and Liu (1996) involves a portion of the leading template strand re-binding the lagging template strand after the leading strand polymerase has passed, but before the lagging strand has been replicated (figure 4.3). It is unclear which endonuclease activities would mediate the cleavage of strands required in this model to generate the formal crossover.

Figure 4.3: A model for RecA-independent mis-pairing between direct repeats.

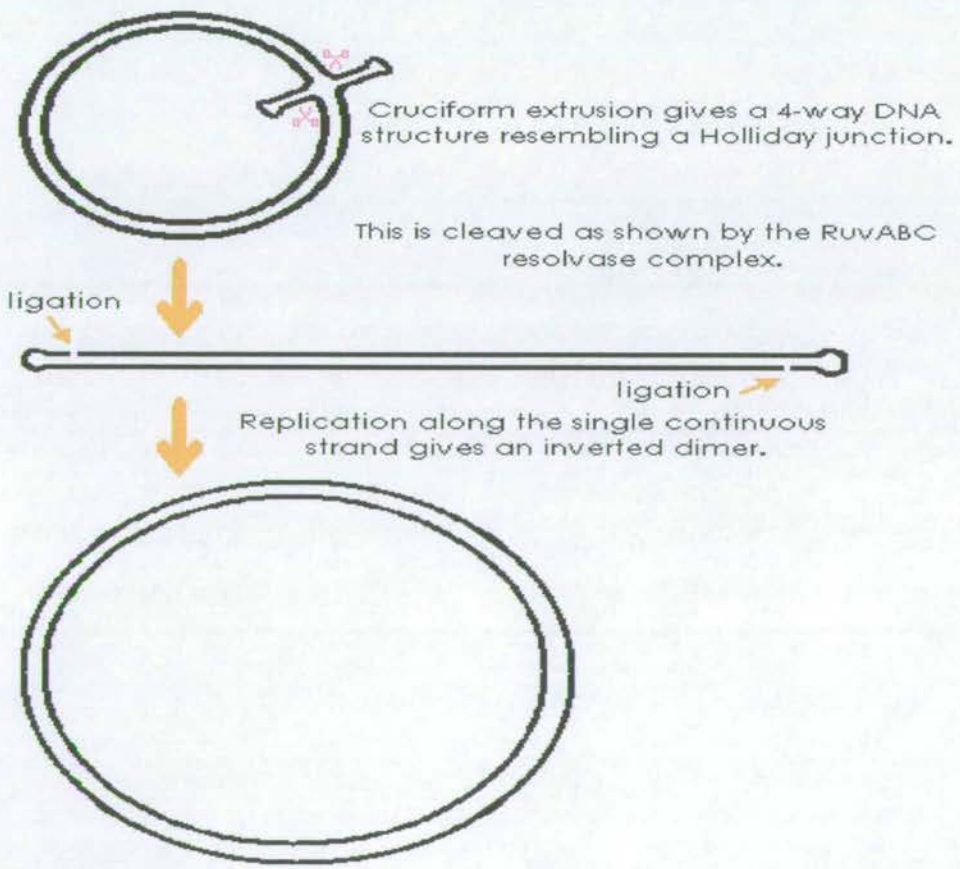


Adapted from Bi and Liu, 1996.

The cruciform/dumbell model for inverted dimer production

An alternative mechanism proposed for plasmids containing direct repeats (Lyu *et al.*, 1999) proposes that a cruciform structure in the plasmid DNA resembles a four way Holliday junction, and is bound by RuvABC (Shiba *et al.*, 1991) and cleaved. This would produce a dumbbell-like intermediate which undergoes replication to give an inverted dimer (figure 4.4). However, inverted dimers are not well maintained within the cell probably because they are effectively giant palindromes, so would be expected to be substrates for SbcCD degradation. Thus recombination models resulting in inverted dimers would be unlikely to be detected using this assay unless occurring at very high frequencies. Although a cruciform made up of a perfect CAG or CTG repeat tract would not contain the preferred RuvC cleavage site of 5'-A/TTTG/C-3', work by Iwasaki and co-workers (1991) suggests that the topology of a four-way DNA junction is sufficient to stimulate cleavage by RuvABC. It is also likely that competition for cleavage at such a site would exist between RuvABC and SbcCD hairpin nuclease. Nevertheless, this model exists as a formal possibility in any sequence capable of forming a cruciform structure.

Figure 4.4: The cruciform/dumbell model for dimer formation.



Methods

Plasmid minipreps were performed on 1ml of LB culture taken from a 5ml overnight inoculated from a single transformant colony. This was duplicated at least six times for each plasmid in each background. 20µl of each miniprep sample was run overnight on an agarose 0.8% TAE gel at 20mV for optimum separation of plasmid forms. Fresh TAE was made up for each gel from 50× stock solution. The gel was stained in the dark using Vistra Green diluted 1:10,000 in TAE buffer. The gel was scanned on a Storm8600 on blue emission spectrum, 100µm pixel size and 900V, as recommended by the Molecular Dynamics manual. Since the blue emission spectrum was being used to measure the relative intensity of DNA bands only 1µl of loading buffer was used for 19µl of sample. Any more loading buffer was found to increase the background intensity of the gel, reducing the accuracy of measurements.

The level of multimerisation was expressed as the proportion of supercoiled dimer molecules present within the total population of the supercoiled monomer and dimer molecules. The supercoiled species were chosen because these two bands are easily identified on native agarose gels amongst bands corresponding to different plasmid multimers and different states of DNA relaxation within each species.

Since plasmid multimerisation occurs in a step-wise fashion, the formation of higher plasmid multimers (other than by rolling circle replication) would occur via the formation of dimers and thus correlate to the number of plasmid dimers present. Indeed, it was found to be the case that strains containing higher plasmid multimers gave high SCD values (see below). Long linear multimers such as those produced by rolling circle replication would not be detected in these experiments because alkaline lysis of cells denatures linear DNA.

Since a dimeric plasmid is capable of intercalating twice as much VG as a monomeric plasmid molecule, the intensity of dimeric bands was halved to give a direct molar comparison of plasmid species. When comparing plasmid dimerisation levels, the average SCD (proportion of supercoiled dimer) value is given.

$$SCD = \frac{\text{supercoiled dimer band intensity} \div 2}{\text{supercoiled monomer band intensity} + (\text{supercoiled dimer band} \div 2)} \times 100$$

The error bars on graph 4.1 show the standard error for each set of values.

$$\text{Standard error} = \frac{\text{standard deviation}}{\sqrt{N}}$$

$$\text{Standard deviation} = \sqrt{\text{Variance}} = \sqrt{\frac{\sum (x - \text{mean})^2}{N - 1}}$$

where N is the number of measurements within a group and x is each individual measurement.

The statistical analysis method chosen to help clarify the significance of values was the arcsin transformation. This transformation normalises the standard deviation so that its magnitude does not depend on the absolute level of dimerisation (since groups with a small amount of dimerisation would be expected to have smaller standard deviations in general). Thus the variation about each mean can be assessed with greater accuracy. The transformation is expressed as:

$$y = \sin^{-1} \{\sqrt{p}\}$$

where p is the measured proportion of dimerisation (SCD).

The problem of multiple pairwise comparisons was overcome using Tukey's honestly significant difference (HSD) test. It makes use of the Student range distribution to identify which population means are significantly different from others. Tukey's test is preferred for conservative ANOVA (analysis of variance) when samples sizes are similar, as it defines a smaller least significant difference

level, making it more powerful. Although Tukey's test results obtained in this chapter were obtained from calculations performed using Microsoft Excel, Tukey's test can be calculated by hand.

Tukey's procedure for a data set which has t sample means and the pooled standard deviation s, based on v degrees of freedom, is as follows:

Two population means i and j are declared different if

$$\left|\bar{y}_i - \bar{y}_j\right| \geq q_\alpha(t,v)\sqrt{\frac{s^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where,

$\bar{y}_i$ is the sample mean of the ith group, n_i is the sample size of the ith group, $q_\alpha(t,v)$ is the upper-tail critical value of the Studentized range for comparing t different populations and α is the family error rate (0.05 was used here to give an overall level of significance of 5%). There is no explicit formula for these critical values but they are tabulated in standard textbooks or books of statistical tables.

Results

The example gel shown below (Figure 4.5) illustrates the increase in plasmid multimerisation attributable to a $(CTG)_{43}$ tract, when compared to pUC18 and pUC18 bearing a $(CAG)_{43}$ tract on the lagging strand, in a *recF* background (DL1368). (p=pUC18, A=CAG_{43}, T=CTG_{43}).

Figure 4.5: **Plasmid CTG43 has an elevated rate of dimerisation in DL1368 compared to pUC18 and CTG43.**

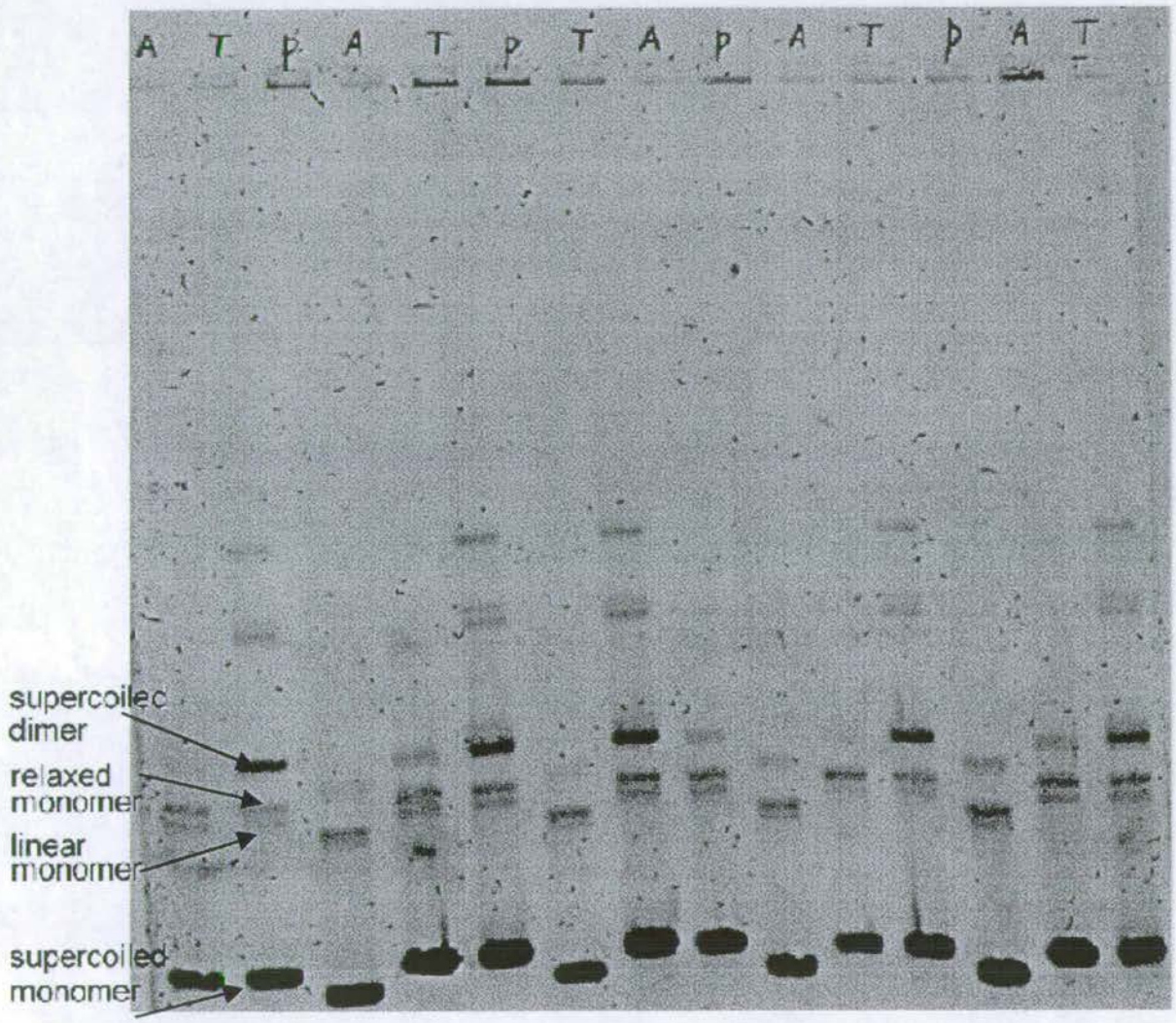


Figure 4.6: The proportion of supercoiled dimer expressed as a percentage of the total monomeric and dimeric supercoiled plasmid DNA obtained from different backgrounds

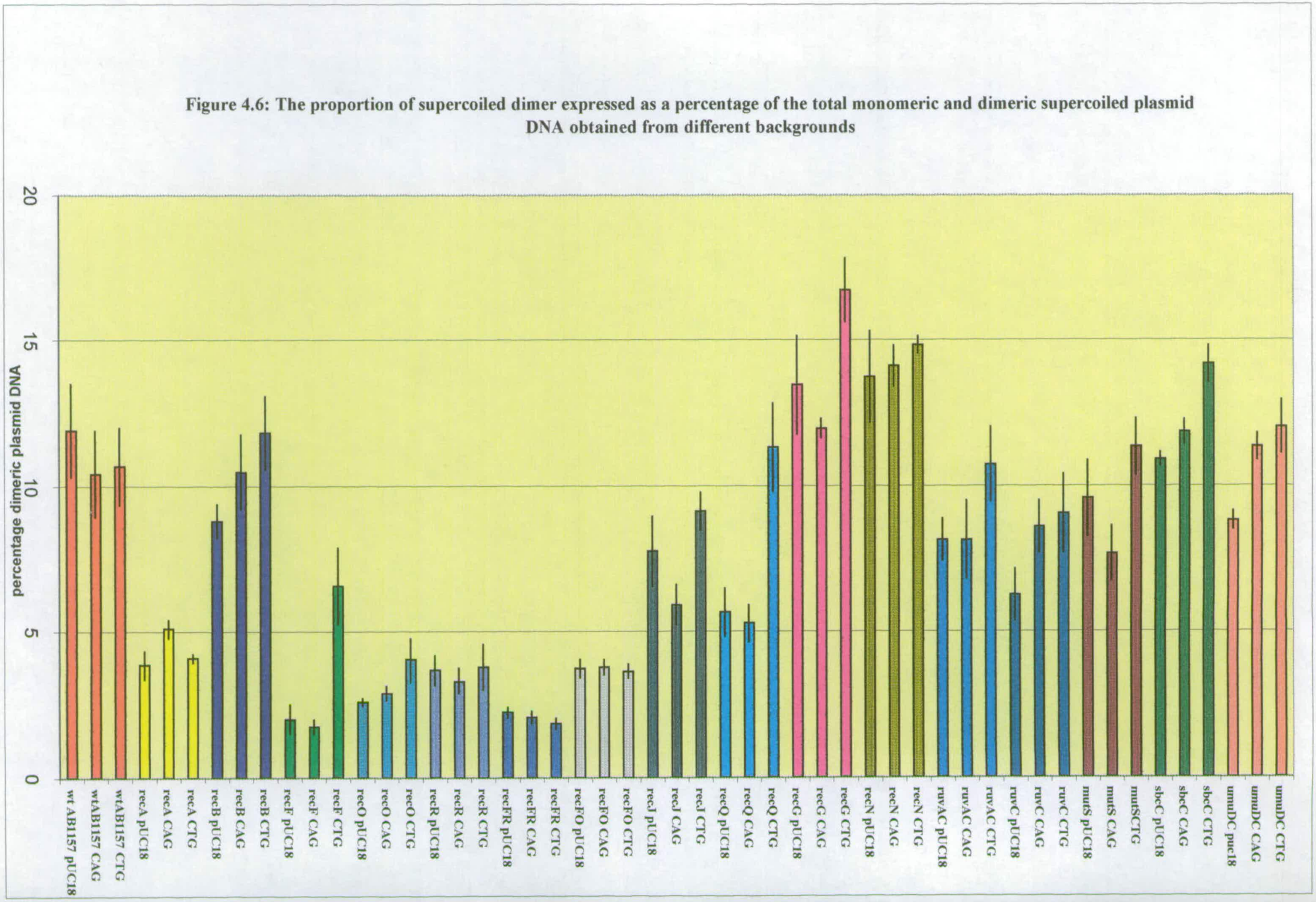


Table 4.1: The mean arcsin values for each TR plasmid tested in each strain.			
strain	Plasmid	mean arcsin	standard error of arcsin
wtAB1157	pUC18	0.349	0.0172
wtAB1157	CAG	0.324	0.0159
wtAB1157	CTG	0.330	0.0159
recA⁻	pUC18	0.196	0.0188
recA⁻	CAG	0.227	0.0188
recA⁻	CTG	0.225	0.0188
recB⁻	pUC18	0.301	0.0201
recB⁻	CAG	0.328	0.0210
recB⁻	CTG	0.350	0.0210
recF⁻	pUC18	0.136	0.0127
recF⁻	CAG	0.130	0.0127
recF⁻	CTG	0.251	0.0133
recO⁻	pUC18	0.156	0.0188
recO⁻	CAG	0.154	0.0188
recO⁻	CTG	0.181	0.0188
recR⁻	pUC18	0.191	0.0188
recR⁻	CAG	0.181	0.0188
recR⁻	CTG	0.190	0.0188
recFO⁻	pUC18	0.194	0.0188
recFO⁻	CAG	0.195	0.0188
recFO⁻	CTG	0.192	0.0188
recFR⁻	pUC18	0.148	0.0210
recFR⁻	CAG	0.142	0.0188
recFR⁻	CTG	0.134	0.0189
recQ⁻	pUC18	0.250	0.0133
recQ⁻	CAG	0.255	0.0149
recQ⁻	CTG	0.326	0.0172
recJ⁻	pUC18	0.269	0.0172
recJ⁻	CAG	0.259	0.0172
recJ⁻	CTG	0.314	0.0172
recN⁻	pUC18	0.376	0.0188
recN⁻	CAG	0.384	0.0188
recN⁻	CTG	0.395	0.0188
sbcC⁻	pUC18	0.336	0.0210
sbcC⁻	CAG	0.351	0.0210
sbcC⁻	CTG	0.385	0.0210
recG⁻	pUC18	0.372	0.0210
recG⁻	CAG	0.353	0.0210
recG⁻	CTG	0.420	0.0210
ruvAC⁻	pUC18	0.287	0.0188
ruvAC⁻	CAG	0.301	0.0172
ruvAC⁻	CTG	0.345	0.0172
ruvC⁻	pUC18	0.249	0.0491
ruvC⁻	CAG	0.296	0.0413
ruvC⁻	CTG	0.301	0.0644
umuDC⁻	pUC18	0.301	0.0188
umuDC⁻	CAG	0.343	0.0188
umuDC⁻	CTG	0.349	0.0172

General levels of pUC18 dimerisation in each strain.

Low levels of pUC18 dimerisation were observed in: *recA*⁻, *recF*⁻, *recO*⁻, *recR*⁻, *recFR*⁻, *recFO*⁻, *recQ*⁻, and *recJ*⁻. Slightly less pUC18 dimerisation was observed in the *ruvC*⁻, *ruvAC*⁻, *mutS*⁻ and *umuDC*⁻ strains compared to wild type, but this reduction was not as pronounced as in the RecF presynaptic pathway mutants listed above. The reduced frequency of overall plasmid multimerisation in *recA*⁻, *recF*⁻, *recO*⁻, *recR*⁻, *recQ*⁻, and *recJ*⁻ was expected as the RecF presynaptic pathway has previously been reported to be the dominant form of plasmid recombination (Kolodner, Fishel, and Howard, 1985). This confirms the sensitivity of this assay.

It would seem that $recA_{13}$ has a slightly less inhibitory effect on plasmid multimerisation than mutations in *recF*⁻, *recO*⁻, or *recR*⁻. It has long been known that the lack of recombination in *recA*⁻ strains severely depresses the level of plasmid multimerisation. In 1974 Hobom and Hogness reported a five to ten-fold reduction of plasmid oligomers in *recA*⁻ strains, and the absence of any multimers larger than dimers. Whilst it was believed that recombination was required for the formation of larger multimers, it was nevertheless acknowledged that a slower, minor pathway of dimer formation was present in *recA*⁻ cells, and was temperature sensitive (Hobom and Hogness, 1974). This low level of highly restricted oligomerization was thought to result from infrequent errors in the replication process leading to sister strand exchange. However, this replicative plasmid dimerisation pathway should not be enhanced in *recF*⁻, *recO*⁻, or *recR*⁻ strains. It is most likely then that the $recA_{13}$ point mutation allele possesses residual activity capable of promoting plasmid dimerisation by recombination.

Although the majority of ss gap repair events result in non-crossovers (Cromie *et al.*, 2000), disrupting the RecF pathway of ss gap repair greatly reduces plasmid dimerisation in this assay. Persistent ss gaps would be converted into DSBs by replication. Gaps on the lagging strand would be converted into lagging strand breaks (biased towards crossover products), where as gaps on the leading strand would be converted into leading strand breaks (biased towards non-crossover products). This may suggest that a very large amount of ss gap repair is occurring in pUC18, so that disruption of ss gap repair can be detected by the reduced generation of crossover

products. Alternatively, it may suggest that the majority of ss gaps found in pUC18 occur on the lagging strand. This latter possibility has important implications to the interpretation of the effect of TR tract orientation on plasmid dimerisation in *recF*⁻ cells (see later).

pUC18 extracted from *recQ*⁻ and *recJ*⁻ cells have significantly higher levels of dimerisation than pUC18 taken from strains containing other mutants in the RecF presynaptic pathway. This suggests that RecA filament is capable of loading onto most substrates present in pUC18 without prior processing by RecQJ. An example of such a substrate would be an ss gap formed by incomplete lagging strand synthesis, or a 5' overhang.

The *ruvAC*⁻ and *ruvC*⁻ strains do not display multimerisation frequencies as low as the strains containing mutations in the presynaptic RecF pathway. This may be because of the ability of RecG to compensate for their missing RuvABC resolvase activities. The elevated dimerisation rates seen in *recG*⁻ may be due to cSDR initiated at stable r-loops elevating linear dimer selection through enhanced replication rates. The very similar arcsin SCD values of pUC18, CAG_{43}, and CTG_{43} plasmids would suggest that cSDR does not occur at significantly elevated frequencies in these TR tracts compared to pUC18.

Dimerisation rates in *recB*⁻ cells were approximately equal to those of wild type cells. This confirms that RecBCD is not required for plasmid multimerisation. Mutations in *sbcCD* might prevent processing of ss overhang into blunt ends, inhibiting RecB-mediated DSBR with its potential for complete degradation of damaged chi-less pUC18. Instead, RecQ and RecJ might be recruited to the break to initiate recombination using the RecF pathway. This diversion of recombinational activity into the RecF pathway is therefore much safer for pUC18 plasmids, and might explain why mutations in *sbcC*⁻ appear to elevate the amount of plasmid dimer above wild type levels.

Figure 4.7: Mean arcsin values for strains in which trinucleotide repeat orientation has no significant effect on plasmid multimerisation

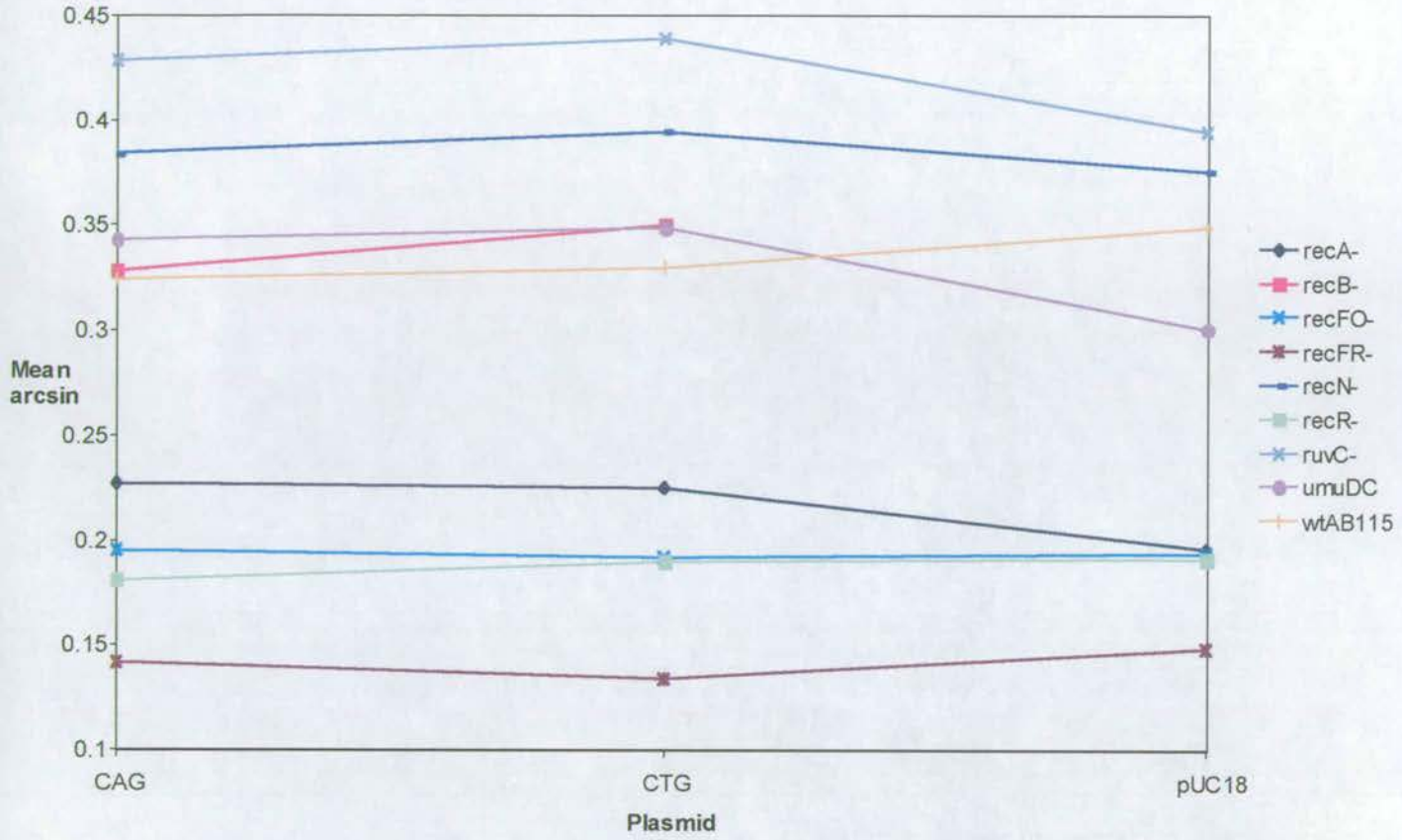
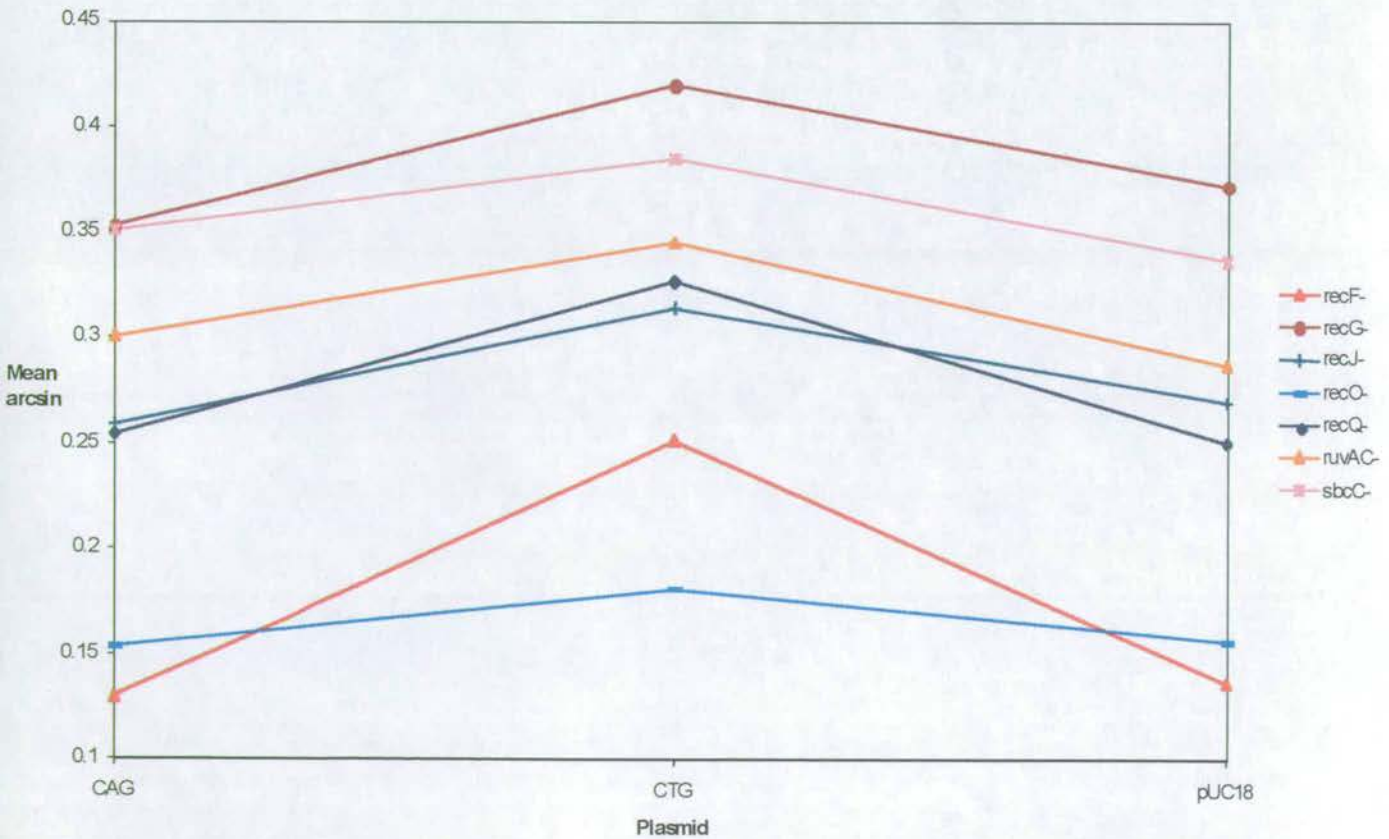


Figure 4.8: Mean arcsin values for strains showing $(CTG)_{43}$-induced plasmid multimerisation.



TR tract orientation dependence on plasmid dimerisation

The presence of TR tracts within pUC18 did not produce higher levels of plasmid dimerisation in wild type cells. This might suggest that in fully recombination proficient *E. coli*, $(CAG/CTG)_{43}$ tracts either do not induce recombination, or induce recombination in a way which does not result in crossover products and subsequent plasmid dimerisation. Recombination mechanisms predicted to result in non-crossover products include ss gap repair, and recombination initiated at replication forks by a leading strand break or at a regressed fork by paired nascent strands. It is also possible that a high background level of recombination in pUC18-based plasmids produced enough crossovers to mask those attributable to $(CAG/CTG)_{43}$ tracts.

No significant differences in dimerisation were observed between the plasmids tested in *recA*$^-$. So, whilst RecA-independent replication crossover might account for a minority of plasmid dimerisation events, this does not occur at significantly higher frequencies in $(CAG/CTG)_{43}$ repeat tracts.

The $(CTG)_{43}$ repeat tract on the lagging strand strongly induced plasmid dimerisation compared to both pUC18 control and $(CAG)_{43}$ in *recF*$^-$ cells ($p=0.002$). Although initial inspection might suggest that similar but less pronounced effects are present in *recJ*$^-$, *recO*$^-$, *ruvAC*$^-$, *recG*$^-$and *sbcCD*$^-$ strains, significant differences were not found between plasmid types using the Tukey's multiple comparison test. Indeed, whilst Tukey's test revealed highly significant differences in plasmid multimerisation between different strains, no significant differences could be demonstrated regarding dimerisation values for the three different plasmids within the same strain (except in *recF*$^-$). In this way, Tukey's test has highlighted the importance of the *recF*$^-$ result, and shown that although significant differences in plasmid maintenance are evident in different strains, each strain maintains a level of plasmid dimerisation which is independent of the presence of a $(CAG/CTG)_{43}$ tract on the plasmid (again, except *recF*$^-$).

Discussion

w.t.

Higher levels of multimerisation are not observed in plasmids bearing TR tracts in wild type cells. This is a significant result. It suggests either that the TR tracts examined here do not induce recombination, or do not produce substrates for recombination leading to crossover products at a level significantly above TR independent crossing over, thus escaping detection by this assay. ss gap repair would produce primarily non-crossover products, as would recombination initiated at replication forks by a leading strand break or paired nascent strands after fork regression.

Results from other strains do display elevated multimerisation in TR plasmids, so TR-induced crossovers can be detected under certain circumstances. TR-dependent dimerisation was detected in certain mutants, probably arising from trapped substrates being processed by a pathway in which HJ resolution is biased towards plasmid dimerisation. Most crossovers would result from ends-in DSBR or recombinational repair of a break in the lagging strand.

RecA

RecA mutations decrease the proficiency of plasmid recombination by 40 - 100 fold (Laban and Cohen, 1981). In the multimerisation experiments carried out here, mostly monomeric plasmid species were extracted from the *recA*$^-$ strain. Although the mean arcsin values of dimerisation for both TR tract plasmids in *recA*$^-$ cells appear to be higher than pUC18, statistical analysis suggests that this is not significant. This eliminates RecA-independent models of dimerisation as mechanisms of TR-induced plasmid multimerisation. The rejected models include RecA-independent sister strand exchange, and the cruciform/dumbbell model of inverted dimer production. A small minority of dimerisation events in all plasmids may occur via RecA-independent mechanisms, such as sister strand exchange, but this appears to occur at approximately the same rate in all plasmids tested. This result would suggest that RecA-mediated homologous recombination determines most CAG/CTG$_{43}$ plasmid dimerisation.

RecF

The reduced level of dimerisation observed in plasmids extracted from cells suffering mutations in the RecF presynaptic pathway would suggest that ss gap repair is the usual method of plasmid dimerisation in *E. coli*. Since only a minority of recombination events mediated by this pathway in wild type cells are expected to result in crossover products, this may suggest that a large amount of recombination occurs in these plasmids which was not detected by the plasmid dimerisation assay used here. Alternatively, it may suggest that the majority of ss gaps found in pUC18 occur on the lagging strand, and when the RecF pathway of ss gap repair is disrupted, replication converts these into lagging strand breaks, which generate crossover recombination products.

In a *recF*⁻ background the decrease in CTG_{43} plasmid multimerisation from wild type is clearly less than that for multimerisation of CAG_{43} and unmodified pUC18. This effect is absent in *recR*⁻ and *recO*⁻. Either it would appear that RecF has a role in preventing plasmid multimerisation at $(CTG)_{43}$ repeats, or recombination at these repeats is independent of RecF. Further analysis provides support for the latter hypothesis. The recombination activity at $(CTG)_{43}$ in *recF*⁻ is lost in the double mutants *recFR*⁻ and *recFO*⁻, suggesting a requirement for the RecOR complex. Since RecF acts to limit RecA filament progression into duplex regions, it is expendable for recombination at DNA ends. Indeed, the absence of RecF may stimulate recombination at DNA ends by removing competition with RecO for formation of RecOR heterodimers (Cox, 2001). Thus a *recF*⁻ genotype may block ss gap repair, but stimulate DSBR by allowing RecOR to use cleaved ends as substrates for RecA. Specifically, RecOR gives better coating of RecA filament onto 5' overhangs by removing the treadmilling bias of RecA which favours filament formation on 3' ends (Dutreix *et al.*, 1991). Therefore, the dependency of CTG_{43} multimerisation on RecOR, but not RecF suggests a DNA end rather than a gap, is the recombinogenic substrate found at $(CTG)_{43}$ tracts. This is consistent with elevated rates of recombination in $(CTG)_{43}$ tracts due to hairpin stalling of replication, giving ss gaps, which are subsequently converted into lagging strand breaks, and undergo RecAOR-mediated recombination to generate dimeric plasmid products.

Formally, there are two more possible explanations for the *recF*⁻ effect. The first is that RecF is simply the most expendable protein for ss gap repair in CTG repeats, as RecA filament progression is limited by other factors specific to CTG repeats, such as DNA secondary structures or branched DNA. The second possibility is that RecF has a function independent of the RecF presynaptic pathway relevant to the maintenance of $(CTG)_{43}$ repeats. In a *priA*⁻ background, *recF*⁻ decreases cell viability and the level of UV-induced SOS de-repression (Sandler *et al.*, 1996). This effect is not observed in *priA*⁻ strains with additional *recO*⁻ or *recR*⁻ mutations. If RecF does have some kind of overlapping activity with PriA, it may be able to prevent plasmid multimerisation brought about by recombination at $(CTG)_{43}$ arrays.

RecQJ

The Tukey test demonstrated that there were no significant differences between plasmid dimerisation in recQ⁻ cells compared to *recJ*⁻ cells. This is consistent with the suggested co-operative interaction between RecQ and RecJ. This co-operative action is an extremely important distinction mechanistically. The co-ordinated action of helicase and exonuclease activities may function in the extension of 3' ends for conventional plasmid recombination, such as in processing blunt ends into 3' overhangs to initiate DSBR. Since TR plasmid dimerisation levels were very similar to those of pUC18 in both *recQ*⁻ and *recJ*⁻ cells, this suggests that recombination in $(CAG/CTG)_{43}$ tracts is independent of processing by RecQJ. Since the RecF presynaptic pathway is the dominant form of plasmid recombination, this may suggest that ss overhangs already exist for RecA loading.

An effect of TR tract orientation on plasmid dimerisation in *recQ*⁻ alone would suggest an helicase activity unwinding DNA secondary structures comparable to that found in larger eukaryotic members of the RecQ helicase family. Persistent secondary structures are expected to form with greater frequency when the CTG-containing strand is on the lagging strand of replication, and if these structures cannot be unwound, they may have to be cleaved. Generation of lagging strand breaks in this way would result in elevated crossover frequency when repaired by DSBR. However, the results presented here do not provide any evidence for RecQ action independent of

RecJ in these repeat tracts. It should be noted that biochemical studies by others (Fry and Loeb, 1999; Kamath-Loeb *et al.*, 2001; Sun *et al.*, 1998; Sun, Bennett, and Maizels, 1999) have only ever shown RecQ-family helicases to unwind secondary structures in (CGG) repeat tracts, so this activity may be restricted to quadruplex DNA, which would not form in $(CAG/CTG)_{43}$ tracts.

RecBCD

There is no effect of $(CAG/CTG)_{43}$ tracts on plasmid recombination levels in the *recB*⁻ strain. If RecBCD were to attempt DSBR of any of these *chi*-less plasmids the entire plasmid would be degraded. An effect on multimerisation would only be expected from this background if TRs induced DSBs, which would be repaired by a less destructive pathway in *recB*⁻ strains. The mean arcsin values for *recB*⁻ do suggest elevated multimerisation for $(CTG)_{43}$ compared to pUC18, but the Tukey multiple comparison test shows that this is not significant. Perhaps the removal of RecF is required to stimulate the safe RecOR-mediated DSBR pathway sufficiently for it to be revealed using this assay. Another possibility in *recB*⁻ cells is that SbcCD still functions in removal of ss overhangs, creating a blunt end normally ideal for RecBCD processing (Thoms and Wackernagel, 1998), but unsuitable for processing by RecFORQJ.

SbcC

The ability of SbcCD to cleave hairpin DNA structures might be expected to elevate the rate of crossing over by producing DSBs. It is therefore perhaps puzzling to note that in *sbcC*⁻ cells, CTG_{43} plasmid dimerisation is not significantly elevated compared to other plasmids (Tukey p=1.0). CTG_{43} plasmid dimerisation in wild type cells is also not significantly elevated compared to CTG_{43} plasmid dimerisation in *sbcC*⁻ cells. This may suggest that either SbcCD does not have a role in hairpin cleavage on the lagging strand, or that this activity is to some extent redundant, and its absence can be compensated for by other activities within the cell capable of ss endonuclease activity at these sites.

RuvAC

Neither the *ruvC*⁻ or *ruvAC*⁻ strains showed significant increases in the dimerisation of CTG_{43}. In addition, the SCD values for both TR plasmids in *ruvAC*⁻ and *ruvC*⁻ strains are very similar to the values in wild type cells. These results casts further doubt on the relevance of the cruciform/dumbbell model of dimer production, and would suggest that cruciform cleavage by the RuvAC complex is not a significant factor in the processing of secondary structures formed in TR repeats. This may be a consequence of cruciform structures not extruding from TR tracts in ds DNA as inverted repeat sequences can, or it may be that any cruciform structures formed might be too transient for cleavage, and would contain mismatched bases every third base, perhaps giving an unusual structure not suitable for RuvABC recognition.

RecG

RecG prevents plasmid replication by unwinding the RNA II primer R-loop at the origin of ColE1 plasmid replication (Vincent and Lloyd, 1996). As such, an increased ColE1 copy number is observed in *recG*⁻ cells in keeping with their constitutive stable DNA replication phenotype (Hong and Kogoma, 1995). Since, *recG*⁻ strains exhibit elevated levels of plasmid replication, they are expected to have a greater selection for dimeric plasmids. This might explain the elevated levels of dimerisation observed in all plasmids extracted from *recG*⁻ cells compared to wild type cells.

Summary

TR tracts do not produce crossovers any more often than pUC18 control in wild type cells. TR-stimulated plasmid dimerisation was observed to be RecA dependent, ruling out spontaneous sister strand exchange and the cruciform/dumbbell models of dimer production. However, the latter model predicts the generation of viability-compromised inverted dimers, which may not occur at a high enough frequency to be detected. TR plasmid dimerisation was also shown to be independent of RuvC, so the cruciform dumbbell model of dimer production proposed for the dimerisation of plasmids containing inverted repeats (Lin *et al.*, 1997) is almost certainly not relevant to dimerisation of the TR plasmids examined here. It is therefore unlikely that TR tracts extrude cruciform structures.

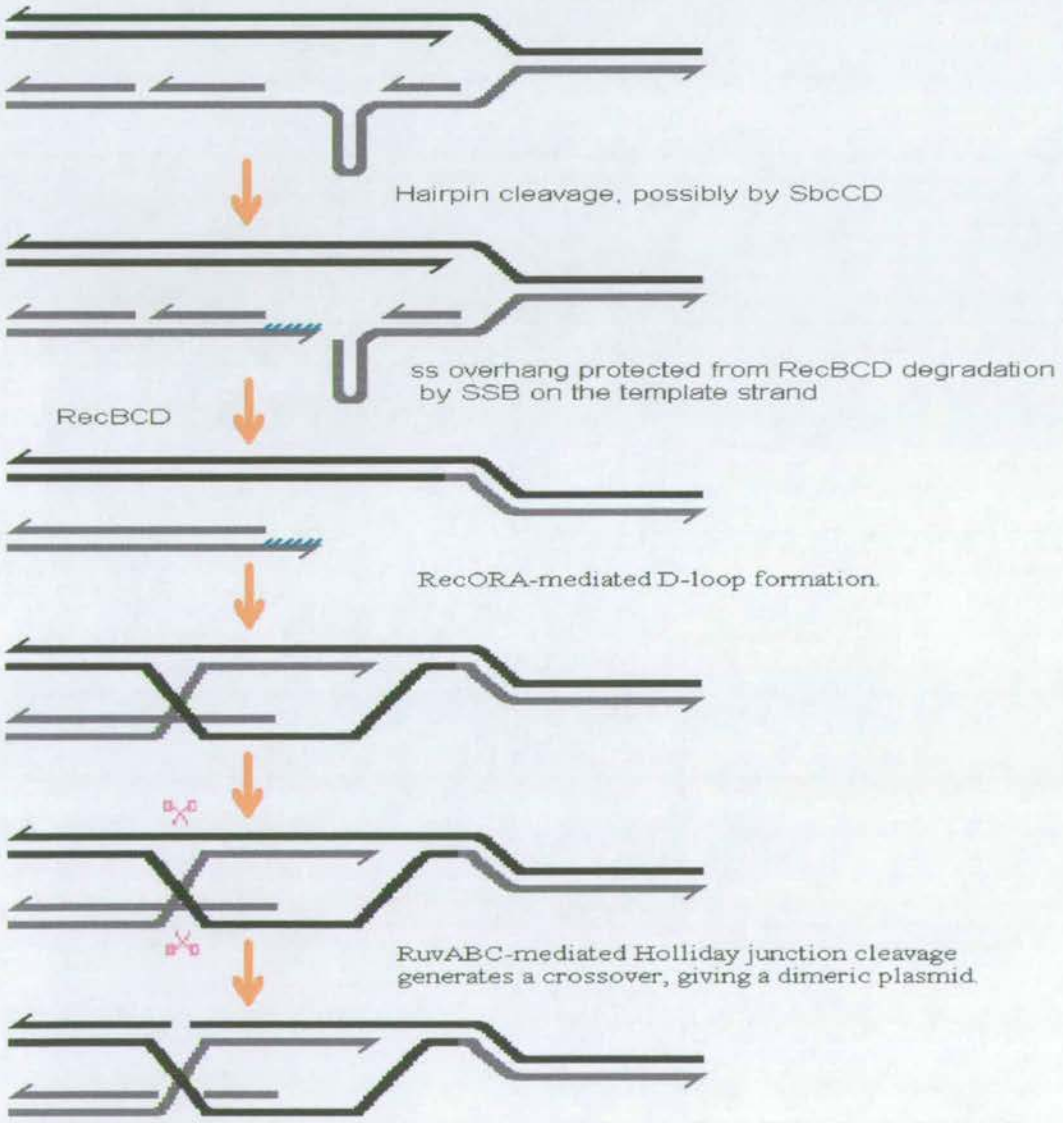
A significant increase in plasmid dimerisation attributable to $(CTG)_{43}$-stimulated recombination was only observed in *recF*⁻ cells, enabling an attempt to define $(CTG)_{43}$ recombination substrates based on the known biochemistry of the absent protein. The observation that a $(CTG)_{43}$ tract stimulates recombination resulting in crossovers in the absence of RecFOR ss gap repair (but not RecOR-mediated RecA loading onto ss overhangs), is informative. It is consistent with an elevated frequency of ss gaps in CTG_{43} tracts (resulting from lagging strand pausing at DNA secondary structures), which are then converted into DSBs either by replication into the gap, or by hairpin cleavage.

It is suggested that processing of DNA ends by SbcCD leads to RecBCD-mediated degradation of the *chi*-less plasmids examined here. An alternative method of presynapsis involving RecOR-promoted RecA filament formation at ss overhangs is thought to be a more productive pathway in these circumstances, allowing DSBR with an accompanying elevated frequency of CTG_{43} plasmid dimerisation. If the RecOR substrate is a cleavage product of SbcCD, one might expect CTG_{43} dimerisation to be reduced in *recF sbcC*⁻ cells. However, there is so much ssDNA present as loops and mismatches in hairpins that lots of other ss exonucleases could substitute for SbcCD cleavage activity. Future work should include transformation of *recD*⁻ cells with the 3

plasmids used in this study, to perhaps give a more direct indication of whether DSBR occurs at elevated frequencies in $(CAG/CTG)_{43}$ tracts.

Hairpin cleavage in an unfinished Okazaki fragment (figure 4.8) might be predicted to be a substrate for the RecFOR presynaptic pathway as the long ss overhang would be expected to be at least partially coated in SSB, preventing degradation by RecBCD. Lagging strand break repair would be expected to give resolvase substrates biased towards generation of crossover products.

Figure 4.9: Hairpin cleavage in an unfinished Okazaki fragment.



Chapter 5: Do CXG trinucleotide repeat tracts induce de-repression of the SOS response?

Aims

TR tracts might be expected to be maintained with some difficulty in a cell if their secondary structures are targets for nucleases or hinder DNA replication. Both of these processes would generate recombinogenic ssDNA, the quantity of which would correlate to the amount of active $RecA^*$ filament within the cell. $RecA^*$ stimulates auto-cleavage of LexA SOS repressor protein, UmuD, and lambda *cI* repressor. Systems have been engineered in which cleavage of LexA and lambda *cI* can be measured by alleviation of transcription repression, allowing production of an enzyme responsible for a specific measurable chemical reaction. Detection of increased rates of these specific reactions have been previously used by others to demonstrate SOS activation in *E. coli* under a variety of stressful and DNA-damaging conditions. Following this rationale, quantification of the degree of SOS response in cells transformed with plasmids containing TR tracts would examine the difficulty the cells are experiencing in maintaining the TR DNA by giving an indication of the requirement of TR tracts for repair by homologous recombination.

TR tracts are not naturally present in *E. coli,* so it is an open question as to whether the cell is already adequately equipped to deal with them, or requires an elevated level of DNA maintenance proteins. If TR tracts do induce the SOS response, this would raise the concept of TRs exerting effects on chromosomal DNA, with important implications for their study in *E. coli* regarding elevated host cell mutation rates.

Causes of SOS de-repression

The SOS system was first characterised by the response of *E. coli* to mutagens in its environment. SOS can be induced by treatment with the DNA-damaging agents 4-nitroquinoline-N-oxide (4-NQO), UV light, mitomycin C (MMC) and methyl methanesulfonates (MMS). All of these mutagens require DNA replication before they induce SOS response, so probably generate signals at broken replication forks. Previous work has also demonstrated the requirement for SOS induction at replication forks stalled by strong protein-DNA interactions. Taki and Horiuchi (1999) showed that the SOS response is de-repressed as a result of replication fork blockage at artificially introduced Tus-Ter complexes in pUC-derived plasmids. Inverted repeats have not been demonstrated to induce the SOS response directly, but cleavage of plasmids by transposase activity at inverted repeats has been shown to induce the SOS response (Shiga *et al.*, 1999).

The SOS response is constitutively activated in certain recombination and replication mutants as the loss of DNA processing activities results in persistent DNA substrates for RecA. SOS induction is observed in *priA*$^-$ strains (Nurse, Zavitz and Marians, 1991; Kuzminov, 1995). *recG*$^-$ strains display a 2 to 3 times greater expression of the SOS response genes than the normal basal level. In *recF*$^-$, *recO*$^-$, or *recR*$^-$ strains, SOS induction is delayed until one round of replication after exposure to harmful levels of UV light, but is then induced to a greater extent than *wt* cells (Hegde, 1995; Whitby, 1995). This may be a consequence of replication forks breaking as they enter ss gaps and result in ds breaks.

Experiments using the *sfiA:lac* SOS reporter system

The first series of experiments in this chapter utilised a *sfiA:lac* reporter system in strain MG1655 (DL1486), an AB1157 background. This strain has been used to demonstrate that the SOS response is de-repressed as a result of replication fork blockage in pUC9-derived plasmids (Taki and Horiuchi, 1999), confirming the sensitivity of this assay in detecting recombination events at stalled forks.

SOS induction was measured in strain MG1655 by a β-galactosidase assay (Figure 5.1). β-galactosidase expression is under the control of a *sfiA* promoter. *sfiA* (suppressor of filamentation) is a gene whose transcription is activated late in the SOS response with the function of inhibiting cell division (Huisman and D'Ari, 1981; Huisman and D'Ari, 1983). Expression from *sfiA* is usually repressed by the SOS repressor protein LexA. Persistent RecA* induces cleavage of LexA, de-repressing the SOS response, and, in the case of MG1655, simultaneous expression of β-galactosidase from the engineered *lac* operon.

Figure 5.1: The principal of regulation of the *sfiA:lac* SOS reporter system. In strain MG1655, LexA autocleavage under conditions which induce the SOS response also result in β-galactosidase expression. Thus measurement of β-galactosidase activity correlates to the level of SOS induction in cells prior to lysis.

Measurement of β-galactosidase activity

A single colony of MG1655 was incubated overnight in 5ml LB at 37°C. Calcium chloride competent cells were made the following day. 50μl aliquots of competent cells were transformed with pUC18 alone, and pUC18 carrying the following trinucleotide repeats: $(CAG)_{43}$, $(CTG)_{43}$, $(CCG)_{24}$, $(CGG)_{24}$, and plated out on LB agar with ampicillin. Following overnight incubation at 37°C, transformant colonies for each plasmid were selected and grown overnight at 37°C in 5ml glucose minimal medium with supplements and ampicillin. The following day, a one in one hundred dilution was performed in minimal medium with supplements and ampicillin, and the culture grown into log phase (OD_{600} of 0.28-0.7) over two and a half hours of shaking at 37°C. At this point, 1ml samples from cultures of untransformed MG1655 and MG1655 transformed with unmodified pUC18, were exposed to varying doses of Mitomycin C (MMC) for thirty min in the dark, to act as positive controls for the sensitivity of MG1655 to SOS induction. MMC is a chemical mutagen known to strongly stimulate the SOS response at a concentration of $0.2\mu gml^{-1}$ LB. Following this treatment, the cultures were washed by pelleting of the cells and resuspension in fresh buffer (see methods in chapter 2 for the safe disposal of MMC). Identical samples from these cultures were left untreated and acted as negative controls showing that this protocol does not induce the SOS response. All cultures were re-suspended in Z-buffer, then underwent lysis and the β-galactosidase assays performed as below.

β-galactosidase activity was measured according to Miller (Miller, 1992). The cells in each 1ml Z-buffer assay mixture were opened by the addition of 2 drops of chloroform and 1 drop of 0.1%SDS, followed by vortexing for 10 seconds. The tubes were incubated at 28°C for 5 min. 0.2ml of ONPG solution was added, and the reaction mixture pipetted into a pre-calibrated 1cm plastic cuvette. The cuvette was placed in a spectrophotometer set to take a "time-scan" measuring the appearance of product at 420nm over 10 min. The resulting printed sheets plotted β-galactosidase reaction velocity and are reproduced below.

Figure 5.2: Beta-galactosidase activity is elevated in MG1655 strains carrying pUC18 plasmids containing trinucleotide repeats

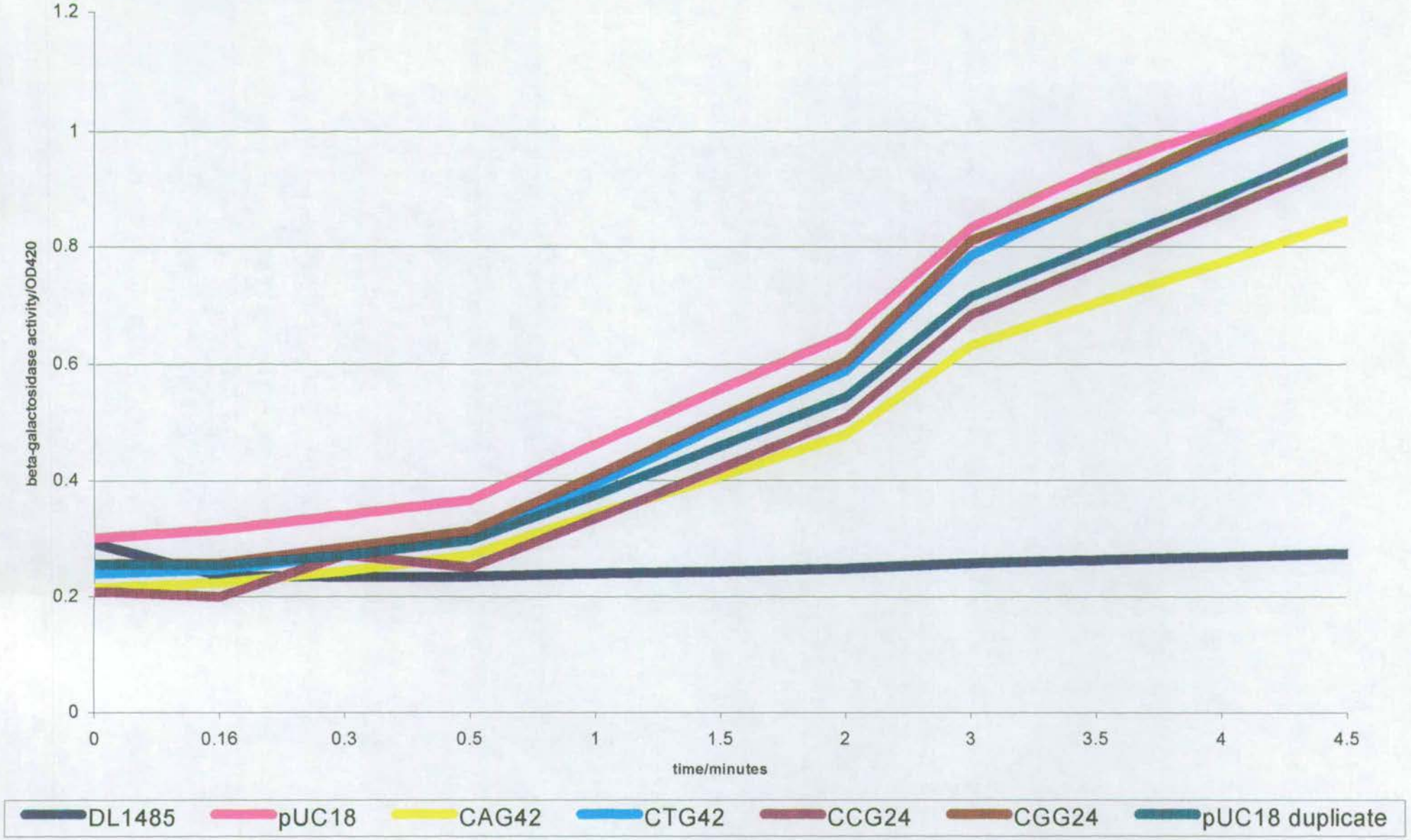


Figure 5.3: Beta-galactosidase activity in MG1655 is minimal in the absence of pUC18-derived plasmids

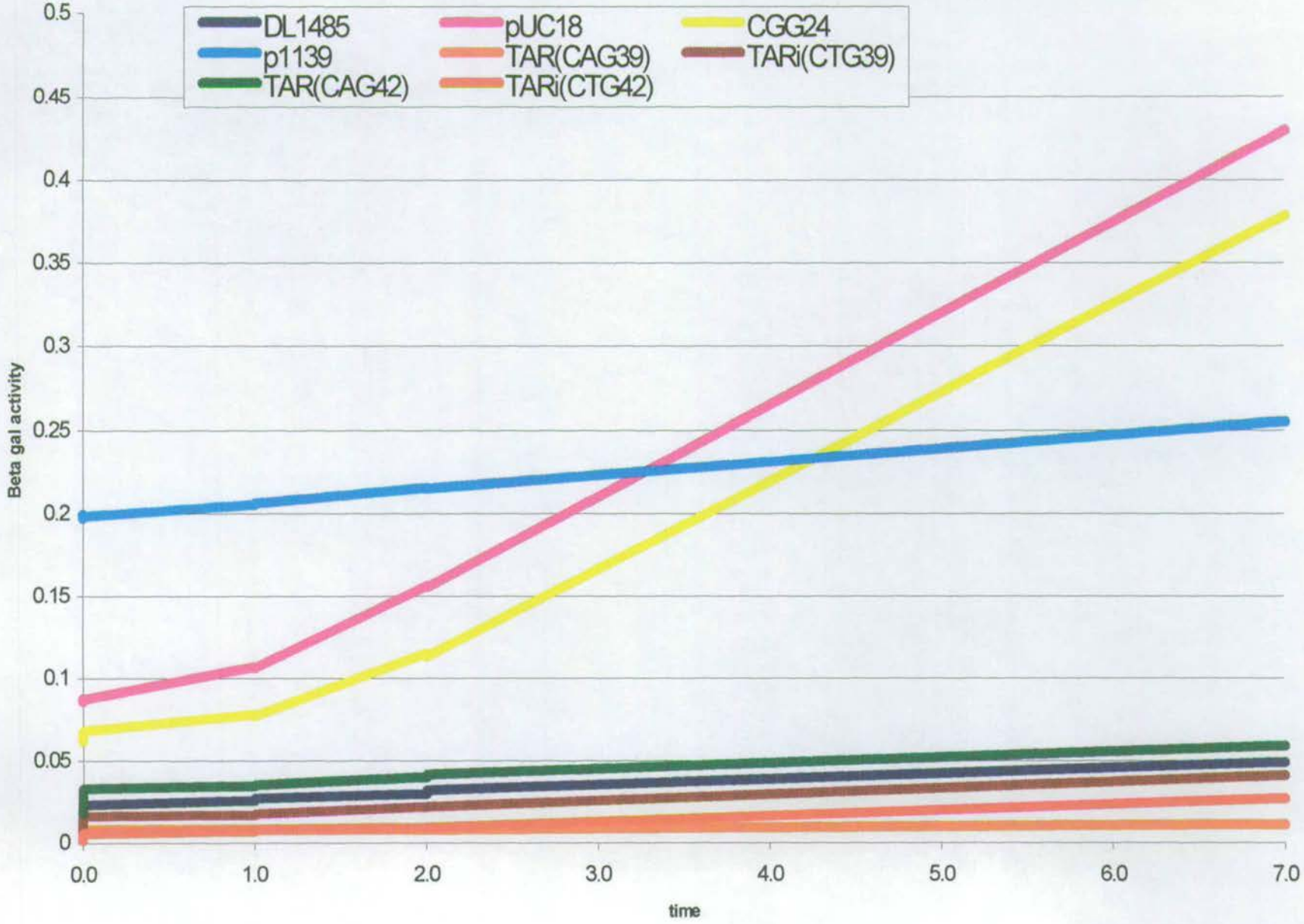
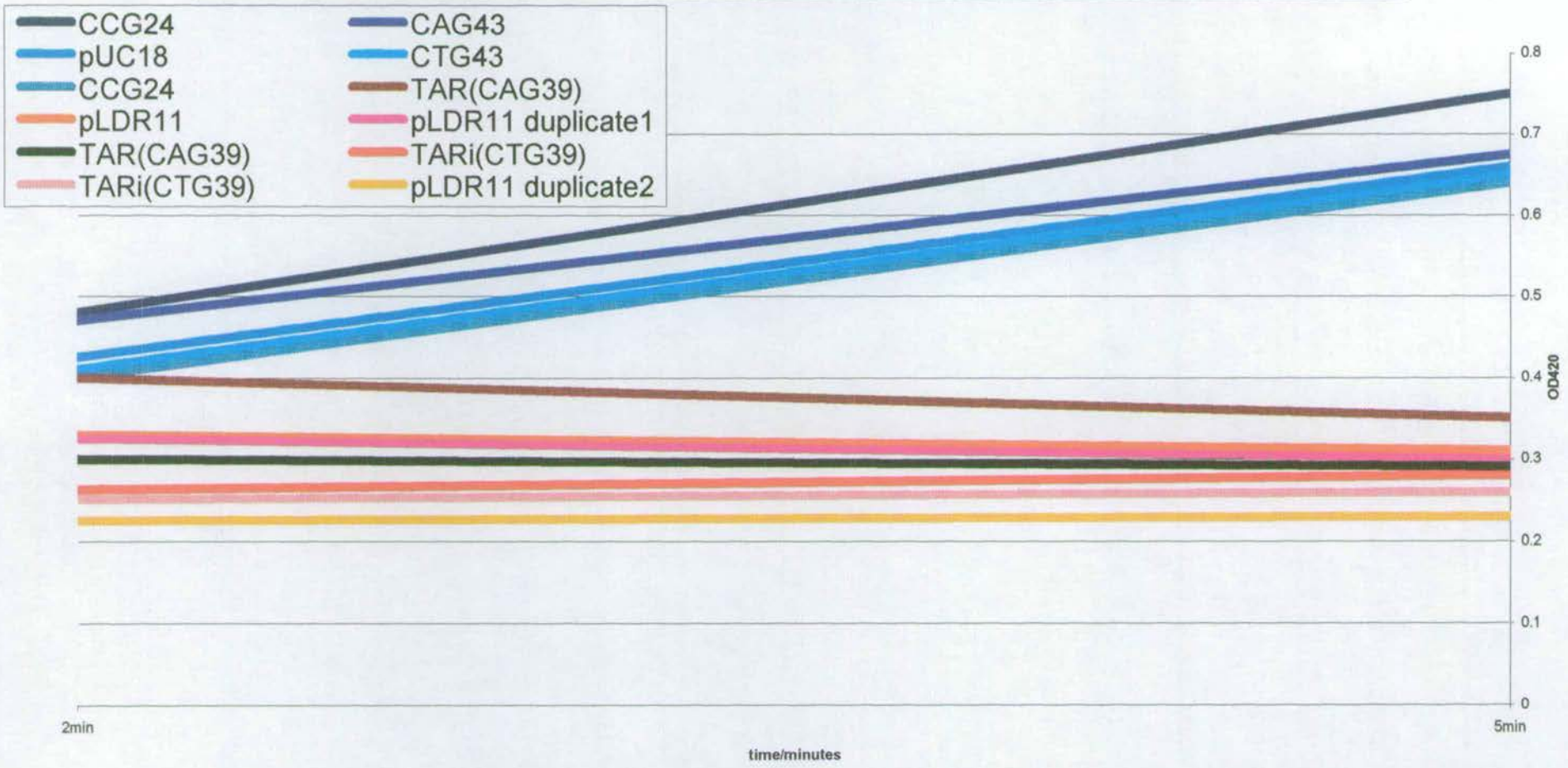


Figure 5.4 TRs present in plasmids not derived from pUC18 do not increase beta-galactosidase activity in MG1655



TAR and TARi are plasmid p1139 carrying $(CAG)_{39}$ and $(CTG)_{39}$ repeats respectively. All other plasmids are derivatives of pUC18.

Figure 5.5: beta-galactosidase expression from the sfiA promoter in MG1655 is proportional to the concentration of mitomycin C used in pre-treatment of cells.

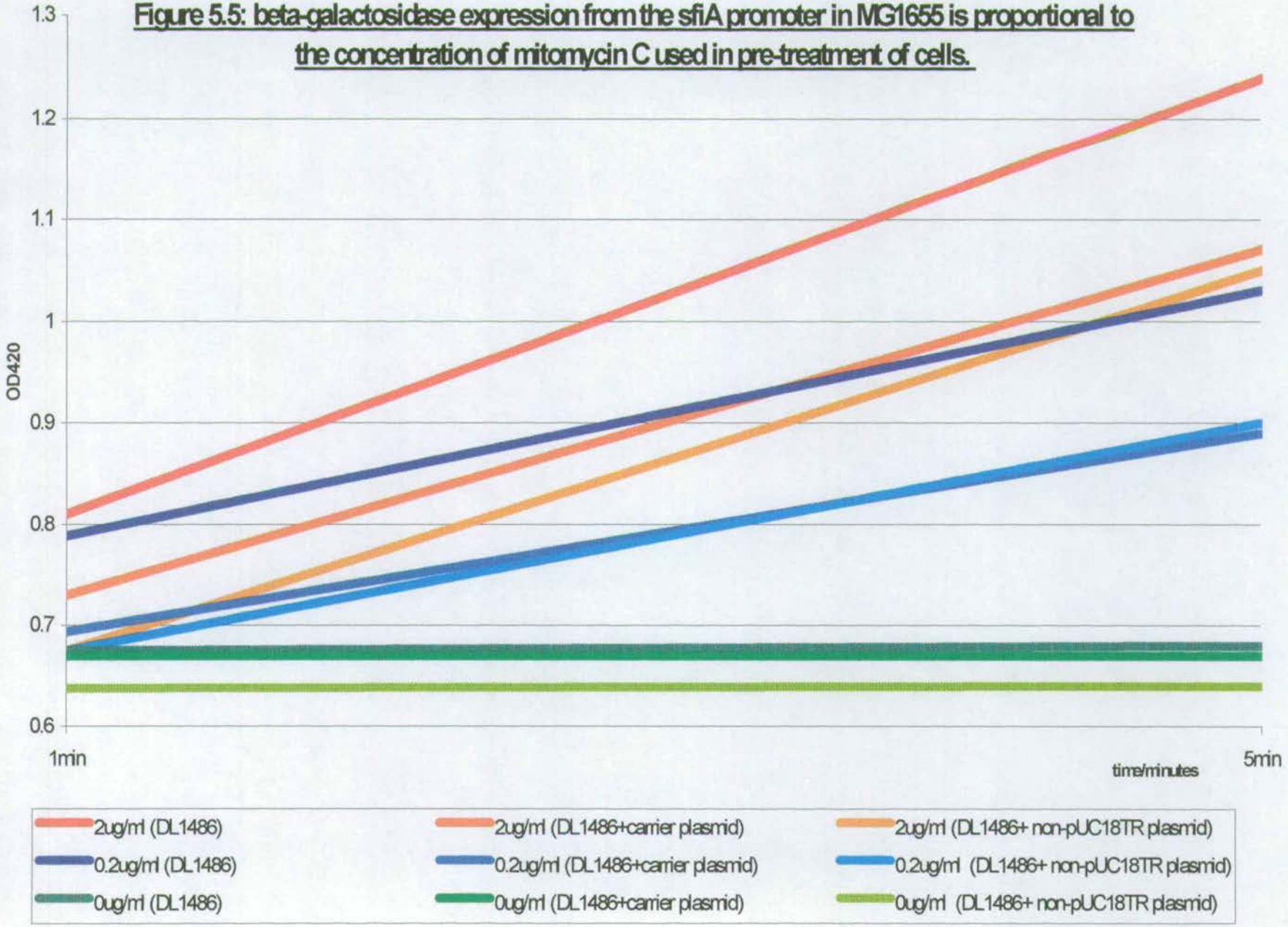


Figure 5.2: Beta-galactosidase activity is elevated in MG1655 strains carrying pUC18 plasmids containing trinucleotide repeats. p148. Freshly transformed cells carrying pUC18 containing trinucleotide repeat tracts were tested for SOS induction using a *sfiA:lac* reporter system by measuring beta-galactosidase activity in log phase cells. Strain DL1485 (MG1655) with and without unmodified pUC18 was used as a negative control.

Figure 5.3: Beta-galactosidase activity in MG1655 is minimal in the absence of pUC18-derived plasmids. p149. Freshly transformed cells carrying pUC18-based plasmids containing trinucleotide repeat tracts were tested for SOS induction using a *sfiA:lac* reporter system by measuring b-galactosidase activity in log phase cells. Negative controls were strain MG1655 (DL1485) with and without either unmodified pUC18 or p1139 (pLDR11). Plasmids TAR and TARi are plasmid p1139 with the addition of a $(CAG/CTG)_{39\text{-}43}$ TR tract.

Figure 5.4: Beta-galactosidase activity in MG1655 is not elevated by p1139 plasmids containing trinucleotide repeats but is elevated by similar plasmids based on pUC18. p150. Freshly transformed cells carrying pUC18-based plasmids containing trinucleotide repeat tracts were tested for SOS induction using a *sfiA:lac* reporter system by measuring b-galactosidase activity in log phase cells. Strain MG1655 (DL1485) with unmodified pUC18 and unmodified pLDR11 was used as a negative control.

Figure 5.5: Beta-galactosidase expression from the sfiA promoter in MG1655 is proportional to the concentration of mitomycin C used in pre-treatment of cells. p151. Strain MG1655 (DL1485) was tested for SOS induction using a *sfiA:lac* reporter system by measuring b-galactosidase activity in log phase cells. Mitomycin C treatment of strain MG1655 at a concentration of 0.2μg/ml LB and 2μg/ml LB was carried out prior to cell lysis. MG1655 that had not been exposed to MMC was used as a negative control for low beta-galactosidase activity. All of the samples were taken from the same log phase culture of MG1655. This experiment served as a positive control for the previous experiments (presented in figure 5.3-5.4), to demonstrate the sensitivity of the *sfiA:lac* reporter system to recombination substrates.

Results and discussion of the β-galactosidase assay experiments

Initial experiments were carried out examining the effect of the presence of trinucleotide repeats within the pUC18 plasmid on induction of the SOS response on MG1655 derived cells during logarithmic phase growth (Fig 5.2). The presence of unmodified pUC18 was not expected to induce the SOS response, but the results showed similarly high β-galactosidase activity in all strains carrying pUC18, regardless of the presence or absence of TR tracts. By contrast the plasmid free control (DL1485) showed negligible activity. Although pUC18 contains only a truncated N-terminus of the β-galactosidase gene, it is thought that some form of genetic complementation has caused this spurious result. The high copy number of pUC18 itself would not induce the SOS response, as this would have been detected in previous SOS-reporter assays in cells containing pUC18 (Lee *et al.*, 2002).

To verify this, further experiments were performed in MG1655 non-pUC18 vectors with and without TRs were used in an identical experiment (results displayed in figures 5.3-5.4). The results presented in figure 5.3 show that β-galactosidase activity is low in DL1485 (MG1655 alone), and MG1655 carrying vector p1139. β-galactosidase activity is only elevated when MG1655 has been transformed with pUC18 or its derivative, CCG_{24}. The presence of a $(CTG)_{39\text{-}43}$ or $(CAG)_{39\text{-}43}$ tract cloned into p1139 does not elevate β-galactosidase activity above that of unmodified p1139. The experiment that provided the results presented in figure 5.4 duplicates this finding. Again, all p1139-derived plasmids gave a low β-galactosidase activity, independent of the presence or absence of TR tracts, and all pUC18-derived plasmids gave a high β-galactosidase activity, also independent of the presence or absence of TR tracts. Figure 5.5 shows the results to a control experiment in which β-galactosidase activity was found to be proportional to the concentration of MMC that MG1655 was exposed to. No MMC resulted in only a residual level of β-galactosidase activity, where as 0.2μg/ml MMC gave some activity, and 2μg/ml MMC gave greater β-galactosidase activity. This demonstrates the sensitivity of the *sfiA:lac* SOS reporter system.

Taken together, these results suggest that the TRs present in pasmids TAR and TARi do not induce the SOS response in *E. coli*. The derivatives of p1139 vector (which does not contain any fragments of the β-galactosidase gene), are referred to as TA because they contains tetracycline and ampicillin resistance genes. TAR is p1139 containing a $(CAG)_{39\text{-}43}$ TR tract in the lagging strand orientation. TARi is p1139 containing a $(CAG)_{39\text{-}43}$ TR tract cloned in the opposite orientation, i.e. $(CTG)_{39\text{-}43}$ on the lagging strand. However, it should be noted that these vectors were low copy number plasmids (approximately 10 copies per cell).

Since plasmids GCC_{24}, and CGG_{24} (carrying the most G-C rich repeats) were not tested, it was decided to confirm and expand upon the above findings by furthering this investigation in a different reporter system. In this way, all of the plasmids taken from the same maxi preps and used throughout the experiments in this thesis could be tested for SOS induction. Thus any comparisons could be made directly and informed by the results to other experiments. This other SOS reporter system, which is also in widespread use, utilises the $RecA^*$ cleavage of lambda repressor protein cI, rather than LexA.

Experiments using the λ-*gal* SOS reporter system

The simultaneous control of error-prone (SOS-induced) DNA repair pathways and lysogenic induction of phage lambda has long been known (Mount, 1977). λ phage entry into a lysogenic or lytic life cycle is determined by the relative levels of two lamda-encoded transcription repressors: cI and Cro. The natural function of the *cI* gene product is complete repression of the lytic growth cycle under conditions favourable for host cell proliferation (Dambly-Chaudiere, 1983). Lysogens are only spontaneously induced in about one in 10^5 cell divisions. However, in situations of SOS induction, cI cleavage warns lambda phage lysogens of possible host cell inviability, and allows progression into the lytic cycle, which encourages phage proliferation and dispersion of their genetic heritage away from the doomed cell.

Figure 5.6: Regulation of the lambda lytic and lysogenic cycles.

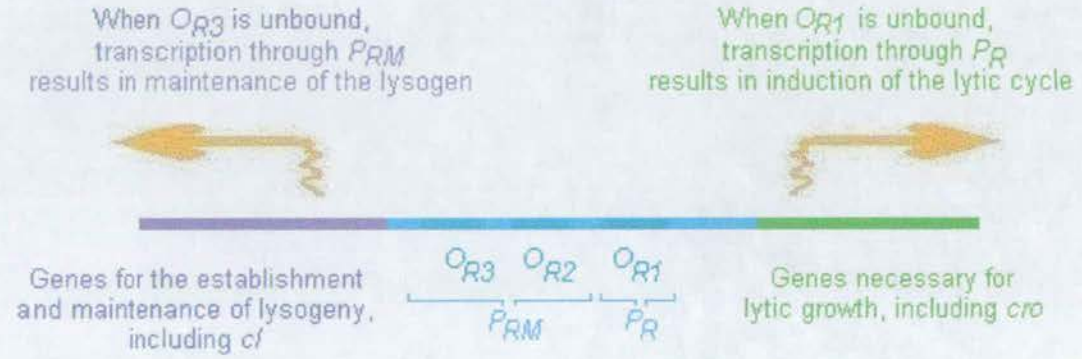
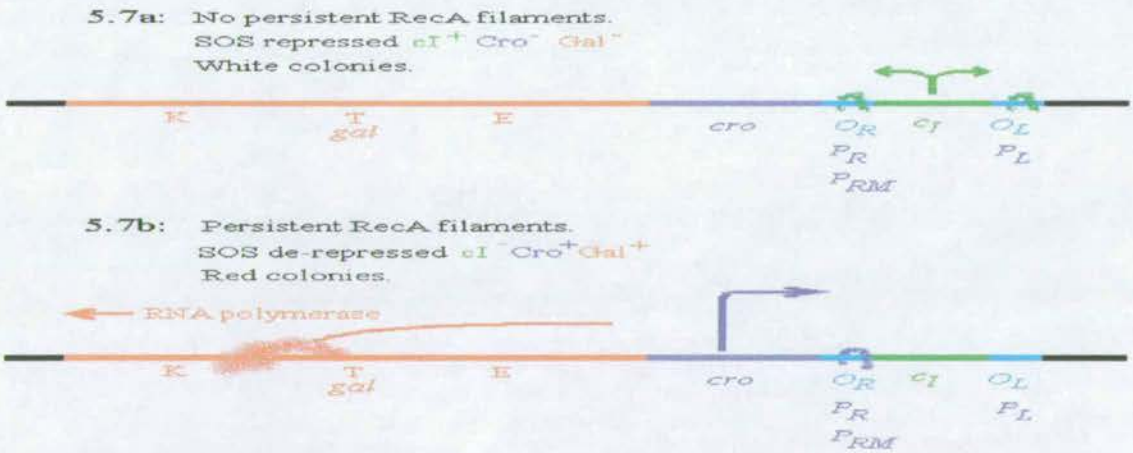


Figure 5.6 shows the three homologous 17bp palindromic subsites making up the p_{RM} and p_R operators. cI has higher affinity for O_{R1} and O_{R2} than O_{R3}, so represses all transcription from the p_R operon, but stimulates its own expression from p_{RM} (cI repressor maintenance) by cooperatively aiding RNA polymerase binding. Cro is an early product of the lytic cycle, and competes with cI for operator binding. The prophage promoters P_{RM} and P_R display different sensitivities to Cro repressor (Meyer *et al.*, 1980), because Cro has a higher affinity for P_{RM} than P_R. Cro inhibits all transcription (including from *cI*) when it occupies the P_{RM} site, so gives positive feedback to the system and commits the phage to at least one generation of lytic growth.

Thus expression of the *pR* operon is an extremely well-regulated genetic switch sensitive to RecA* activity. The same promoters have been engineered to regulate galactokinase expression in strain DL1437 in the form of a *λ-gal* fusion (Toman *et al.*, 1985). This system is a heritable reciprocal repression between phage λ *cI* and *cro* (figure 5.7). Transcription of the *gal* operon is under the control of the *λpR* promoter from a defective prophage, so that non-SOS-induced cells will synthesise λcI repressor, giving white colonies corresponding to a Gal⁻ phenotype. ssDNA-RecA* filament directly stimulates proteolytic auto-cleavage of cI in the same cellular conditions as SOS induction, allowing transcription of the *gal* operon and the *cro* gene, producing red colonies. In nature, lambda lysogen host cell immunity to superinfection (provided by cI repressor) is lost permanently after transient derepression of the lytic cycle because Cro protein can repress *cI* expression. Since this assay is controlled by the same genetic switch, even a temporary SOS induction would give a Gal⁺ phenotype lasting for many generations, making this an extremely sensitive assay even for transient SOS induction.

Figure 5.7: Principles of the galactokinase assay for the measurement of SOS induction in DL1437.



Methods for the *λ-gal* SOS reporter assay

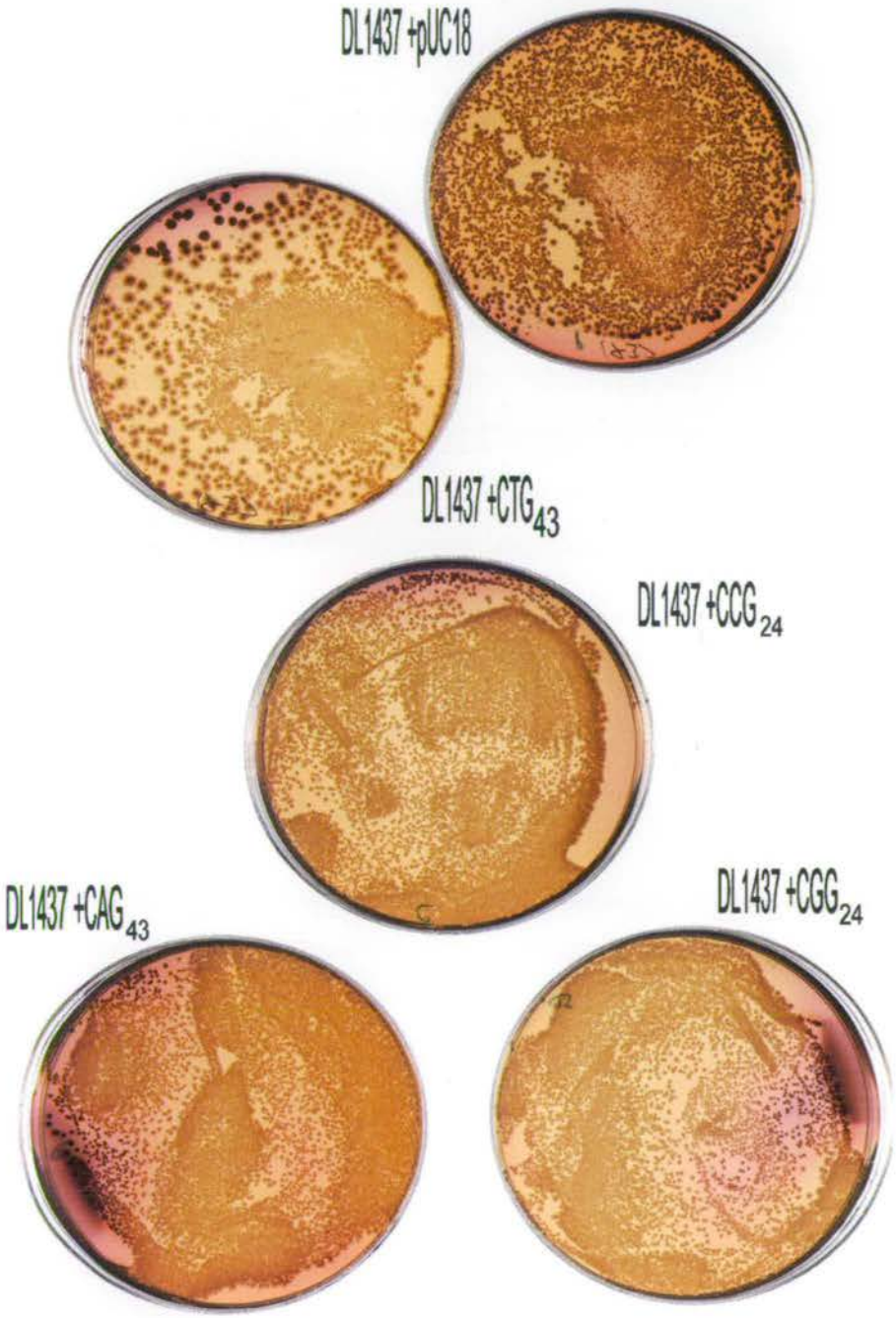
A single colony of MT101 (DL1437) was incubated overnight in 5ml LB at 37°C, and calcium chloride competent cells made. 50µl aliquots of competent cells from the same strain were transformed with pUC18 alone, and pUC18 carrying the following trinucleotide repeats: $(CTG)_{43}$, $(CAG)_{43}$, $(GCC)_{24}$, $(CGG)_{24}$, and plated on LB agar with ampicillin. Following overnight incubation at 37°C, transformant colonies for each plasmid were selected and grown overnight at 37°C in 5ml LB. Serial dilutions of this broth were then plated out on MacConkey medium to guarantee appropriate cell densities for reporter plates. After plating on MacConkey plates, no longer than 48 hours incubation at 37°C was permitted, since SOS induction occurs in stationary phase cultures (Taddei *et al.*, 1995). Plate images were taken by scanning on a flat bed scanner.

MT112 (DL1526), a *recA1* derivative of MT101 was used as a negative control to demonstrate white colony colour. The RecA mutation makes the strain unable to stimulate auto-cleavage of either LexA (so cannot induce the SOS response), or lambda cI repressor protein (so cannot allow *gal* expression). A mitomycin C incubation (protocol identical to that used in the β-galactosidase assay above) was used as a control for the appearance of SOS-induced red colonies.

Results of the *λ-gal* SOS reporter assay in MT101.

The first plate shown on this page demonstrates that pUC18 does not induce cI repressor autocleavage. None of the pUC18 derivatives bearing $(CAG)_{43}$, $(CTG)_{43}$, $(CCG)_{24}$, or $(CGG)_{24}$ induced cI repressor autocleavage. Residual areas of red on the MacConkey plates are due to low colony densities. The colonies themselves are all white.

Figure 5.8: None of the pUC18 derivatives induced cI repressor autocleavage.



The plate shown below (figure 5.9) is an example of DL1437 incubated with MMC to act as an SOS-induced control for red colony appearance.

Figure 5.9: The addition of MMC to DL1437 carrying pUC18 does induce cI repressor autocleavage.

Since pUC18 does not contain a *chi* site, it is possible that the sensitivity of these experiments were compromised by RecBCD (ExoV) activity. DSBs occurring in any of these plasmids would be a target for RecBCD loading, with the potential for subsequent unhindered digestion of the entire plasmid. Digestion of dsDNA by RecBCD is highly processive, and would be unlikely to provide opportunity for $RecA^{*}$ loading. Inclusion of a *chi* site within the plasmid would not provide an unequivocal solution to this problem, because *chi* sites only switch off the exonucleolitic activity of RecBCD in 25-40% percent of encounters (Taylor and Smith, 1992; Dixon *et al.*, 1993).

The *λ-gal* SOS reporter assay in *recD*⁻ cells.

One way to ensure that potentially recombinogenic plasmids are not being degraded during DSBR due to the absence of *chi* sites would be to duplicate the experment in a *recD*⁻ SOS reporter strain. With this aim, a P1 lysate was prepared in DL1554 containing the *recD*$_{1901}$*::Tn10* allele. This was transduced into the *λ-gal* SOS reporter strain (DL1437), and recombinants selected on tetracycline L-agar, before being carried forward for genetic confirmation of a *recD*⁻ phenotype.

The lambda spot test is a good genetic test for a *recD*⁻ phenotype. Lamda *red*⁻ *gam*⁻ phage carrying a *chi* site (DRL105) grow far better on *recD*⁺ cells than phage without *chi* sites (DRL104). Both lambda strains have good growth on *recD*⁻, since there is no ExoV activity. Thus the loss of DRL105 growth advantage compared to DRL104 indicates a *recD*⁻ phenotype.

Unfortunately the usual method of checking a transduced *recD*⁻ allele using a lamda spot test cannot be used directly in DL1437-derived strains, because the expression of cI prevents further lambda infection. However, successful *recD*⁻ transduction can be verified by a back-cross. P1 phage were grown on candidate transformant single colonies, then introduced into AB1157 wild type cells (DL513); again, recombinants were selected for on tetracycline L-agar (subsequently labelled strain DL1556). A lambda spot test was performed on these colonies with the knowledge that a *recD*⁻ phenotype could only occur as a result of P1 transduction from DL1555, itself a transductant from DL1554. A *recD*⁻ phenotype was indeed observed as equally large plaques were formed by both DRL104 and DRL105 i.e, independent of *chi*. In this way, the confirmed re-introduction of the *recD*$_{1901}$*::Tn10* allele proved the existence of this allele in the intermediate strain DL1555.

recD⁻ strains are hyper-recombinogenic since RecA* formation still occurs on the unwound 3' ssDNA, but the absence of ExoV activity prevents extensive *chi*-less DNA degradation. MMC was used to confirm that the sensitivity of the Gal assay was not altered by the absence of RecD exonuclease. A negative control of DL1555 alone was included to verify that the *recD*⁻ allele did not induce the SOS response.

Results to λ-*gal* SOS reporter assay in *recD*⁻ cells (DL1555)

None of the plates shown below (figure 5.10) demonstrate induction of the SOS response, regardless of the presence or absence of large numbers of TR tracts. It is important to note that induction of the $recD_{1901}$::*Tn10* allele into DL1437 did not result in SOS induction, despite the hyper-rec phenotype observed in *recD*⁻ strains.

Figure 5.10: None of the pUC18 derivatives induced cI repressor autocleavage in DL1555.

DL1555+CGG_{24}

DL1555+pUC18

DL1555+CTG_{43}

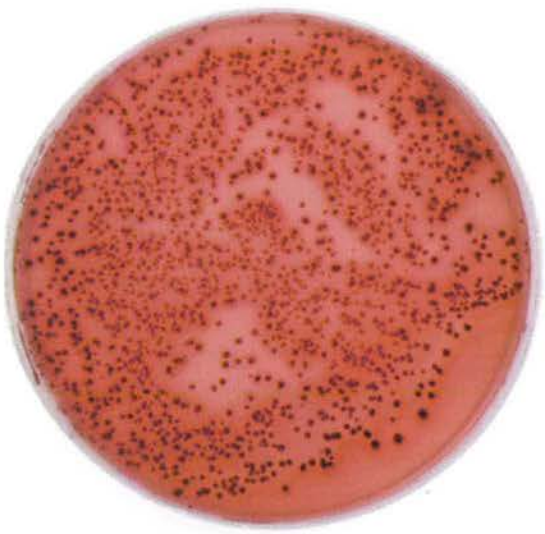
DL1555+$(CCG)_{24}$

DL1555+$(CAG)_{43}$

The two positive control plates shown below demonstrate that DL1555 displays sensitivity to SOS-inducing conditions.

Figure 5.11: The addition of MMC to DL1555 does induce cI repressor autocleavage.

DL1555+MMC



DL1555+pUC18+MMC

Unified discussion of *sfiA:lac* and λ gal SOS reporter assays

The results to the experiments performed in this chapter using the β-galactosidase assay in strain MG1655 found elevated β-galactosidase activity in cells carrying any pUC18-derrived plasmid, regardless of the presence or absence of TR tracts. This may have been a consequence of genetic complementation with the residual 5'-terminal fragment of the *lacZ* gene present on pUC18, as MG1655 carrying plasmids derived from vector p1139 did not display elevated levels of β-galactosidase activity. Indeed, derivatives of vector p1139 carrying $(CAG)_{39}$ or $(CTG)_{39}$ displayed β-galactosidase activity as low as that of unmodified p1139. The *sfiA:lac* SOS reporter system of MG1655 was found to be sensitive to RecA filament formation brought about by treatment of cultures with the SOS-inducing mutagen, MMC. However, whilst these experiments showed that $(CAG)_{39}$ or $(CTG)_{39}$ tracts present on the low-copy number plasmid p1139 do not induce the SOS response in *E. coli*, it was decided to use a second SOS-reporter system to examine the high copy number pUC18 derived plasmids (CAG_{43}, CTG_{43}, CCG_{24}, and CGG_{24}) used in the rest of the experiments presented in this thesis.

The second set of experiments presented in this chapter to measure the level of SOS induction used the λ gal reporter system in strain DL1437. Galactokinase activity was elevated (giving red colonies on MacConkey medium) subsequent to treatment of DL1437 with MMC, proving the sensitivity of the assay. In contrast, galactokinase activity was found to be minimal (white colonies on MacConkey medium) in DL1437 alone and DL1437 carrying pUC18 with $(CAG)_{43}$, $(CTG)_{43}$, $(CCG)_{24}$, or $(CGG)_{24}$ tracts. This shows that the SOS DNA damage response operon is not de-repressed in *E. coli* carrying several hundred TR tracts of clinically relevant length.

During the course of the experiments using either of these assays, a supplemental observation was made that no differences in growth rates were measured during log phase growth from overnight cultures of cells carrying the different plasmids tested.

This shows that log phase growth was not inhibited in cells carrying TR plasmids, further suggesting that the SOS response is not induced.

Both the *sfiA:lac* and λ gal SOS reporter assays used in this chapter were functional, sensitive and have previously been employed in other research examining causes for SOS induction in *E. coli*. The high copy numbers of the pUC18-derrived plasmids used in the λ gal reporter experiments should have provided the maximum opportunity for persistent RecA* filament formation at any recombination substrates formed within the TR tracts. The fact that none was detected is itself interesting. The most fundamental conclusion is that none of the structures formed during the replication or repair of these TR tracts are able to act as a signal for serious DNA damage, requiring the postponement of cell division and up-regulation of the cell's capacity to repair DNA. Structures present at TR tracts might have induced enough LexA cleavage to induced the SOS response if they are persistent substrates for RecA* filament, or if they require frequent recombinational repair (with the consequence of increased RecA* activity in the host cells).

It may be that the TR tracts used in this study are too short to form secondary structures stable enough to stall replication forks or necessitate cleavage by endonucleases, so they do not become frequent or persistent targets for RecA* filament formation. Many copies of TR tracts of these lengths can be comfortably maintained (within the confines of the instability observed in chapter 7) by normal levels of recombinational repair available to *E. coli*.

However, results from multimerisation (chapter4) and competition studies (chapter6) in this thesis suggest that the TR tracts examined here do form recombination substrates under some circumstances, and do require some degree of maintenance. However, there may be quite a tolerance of recombinational activity before SOS induction in the *λ-gal* assay, because the promoter used in this reporter assay normally regulates expression of the lytic growth genes, which are de-repressed relatively "late" in the SOS response. The assays would be more sensitive if early SOS response gene promoters were used. However, both of these systems are

acknowledged as reliable methods for measuring SOS induction, and their sensitivities were measured in this study in parallel control experiments. Clearly large amounts of recombination or persistent RecA* filament would have been detected. Once again, the high copy number of the TR plasmids within cells have improved the sensitivity of these assays by providing many opportunities for RecA* formation along TR tracts. Because some recombination substrates might be expected to be processed more quickly than others, and some blocked replication forks do not require RecA activity for re-initiation, the discussion which follows attempts to discriminate between substrates for DNA repair which are more or less likely to result in detectable RecA* filament activity.

It is clear from these negative SOS results that any recombination carried out in these TR tracts is processed rapidly, as RecA* filament is never persistent enough to induce the SOS response. RecA* filament might be expected to be relatively persistent if it was required to search for regions of homology on other copies of the plasmid, such as in inter-plasmid recombinational repair of ds gaps. One obvious requirement for this would be in the repair of DSBs caused by cruciform cleavage in dsDNA or hairpin cleavage in ssDNA. Also, hairpin cleavage on one strand would generate a nick, which would produce a DSB when encountered by a replication fork. Extensive DSBR would require large amounts of RecA* filament, but the sensitivity of the assays used here to DSBR was doubtful because of the lack of *chi* sites on plasmids resulting in complete degradation rather than RecA loading. The same experiments were performed in *recD*⁻ cells proficient in RecA loading at DSBs and not capable of plasmid degradation, but again, gave negative results for RecA* filament activity.

RecA* filament might be more short-lived when mediating strand exchange between duplexes held in close proximity, such as intra-plasmidic recombination between daughter strands behind replication forks, or in replication fork re-initiation. Because intra-plasmidic recombination would be expected to require less "searching" time by RecA* filament to find regions of homology, recombination at stalled and broken replication forks will be considered further.

Replication forks are "naturally" stalled at *ter* (replication terminator) sequences (Bidnenko, Ehrlich, and Michel, 2002). Back to back replication terminators engineered into the terminal region of the *E. coli* chromosome only induced the SOS response after arrest of at least two attempted rounds of replication (Sharma and Hill, 1995). This suggests that stalling of the first fork by a "natural" block to DnaB replicative helicase is not processed by the cell's DNA repair activities. It is the collapse of the second replication fork at the first stalled fork which results in free DNA ends. Since this was not observed in the experiments carried out here, it is possible to say that even if secondary structures in TR tracts do pause replication forks in a manner similar to Tus-*ter* replication traps, they are processed quickly enough that they do not pause two consecutive forks to generate ds ends.

"Un-natural" replication fork pausing and breakage due to mutations in factors associated with replication fork processivity, undergo RuvAB-mediated regression and processing by RecBCD or RuvABC (Flores *et al.,* 2001). Activities present in the RecF pathway can also mediate repair of this type of arrested fork (Bidenko *et al.,* 1999). Replication forks stalled or broken at UV lesions may require RecQJ (Courcele and Hanawalt, 1999) or RecF activities for re-initiation (Courcelle *et al.,* 1997 and 1999). All of these "un-natural" blocks to replication require processing in a fashion which might induce the SOS response if occurring at high frequencies. In addition, *in vitro* work by Robu, Inman and Cox (2001), indicate that direct loading of RecA alone promotes replication fork regression.

Stalled forks resulting from obstruction of a polymerase on one strand would result in uncoupling of the polymerases, and generation of a ss gap, suitable for RecA* filament formation. Inhibition of DNA synthesis on one strand, followed by re-initiation as the other polymerase carries the fork forward would also result in the formation of ss gaps suitable for RecA* filament formation. The results to the experiments presented in this chapter would suggest that if replication forks do stall as a result of polymerase uncoupling during passage through TR tracts, this does not frequently result in persistent ss gaps.

Alternatively, if a DNA secondary structure occupies both template strands, the polymerases should stall together, leaving no DNA ends or ss gaps for RecA* filament formation. Fork regression away from the structure would lead to nascent strand annealing (giving a ds end), but this could be degraded by RecBCD without the need for RecA* filament formation, or threatening degradation of the entire plasmid. Cleavage of the chickenfoot Holliday junction by RuvABC would generate recombinogenic ds ends, but this may occur far less frequently than nascent strand degradation by RecBCD, as this reaction was observed in *recB*⁻ mutants (Flores *et al.*, 2001). So it may be that stalling of a replication fork by a stable DNA secondary structure resembles stalling by a very strong protein-DNA interaction than by physical defects in the template DNA duplex or replisome processivity. In this respect, replication fork pausing (rather than collapse) and maintenance of the intact replisome (probably favouring fork regression rather than re-initiation by recombination) might be possible in TR tracts.

Fork regression would give opportunity for processing of the secondary structure blocking replication of the TR tract. Hairpin cleavage and recombinational repair similar to palindrome processing in *E. coli* is not the only method of resolution. Recent work has shown that RecQ and its eukaryotic homologs are able to unwind tetrahelical DNA, alleviating inhibition of DNA polymerase through $(CGG)_n$ repeat tracts. In addition, toposimerases may be able to push the hairpin-duplex equilibrium in favour of duplex formation by reducing the negative superhelicity favouring cruciform or hairpin extrusion. Neither of these processes would require RecA, but it is likely that both would require fork regression.

Conclusions

The aim of this series of experiments was to determine whether TR tracts stimulate RecA-dependent recombination, and whether recombination substrates are persistent or abundant enough to de-repress the SOS response. No induction of the SOS response was detected in TR tract-containing cells as detected by either the *sfiA:lac* or the λ gal assays. The assays used in these experiments were sensitive to SOS induction using MMC, and were the same promoter-reporter gene fusions used by others to demonstrate SOS induction at stalled replication forks. Therefore, if recombinational repair is required for processing of hairpin structures or their cleavage products, the intrinsic level of repair present in un-induced *E. coli* is sufficient to mediate this process, even when the TR substrates are present on high copy number plasmids.

If secondary structures are formed frequently within TR tracts and do block replication, the results presented here indicate that they probably form structures that are not reversed by $RecA^{*}$. Fork reversal by RecA has previously been inferred from genetic evidence at stalled forks in $dnaB^{ts}$ mutants (Seigneur, Ehrlich, and Michel, 2000), and *in vitro* studies using structures designed to resemble leading strand gaps at forks collapsed by polymerase uncoupling at a leading strand lesion (Robu, Inman, and Cox, 2001). Stalled forks shown not to require RecA for reversal include those found in rep^{-} (Michel *et al.*, 1997), $holD^{-}$ (Flores *et al.*, 2001), $dnaN^{ts}$ and $dnaE^{ts}$ strains (Grompone *et al.*, 2002). Therefore, if forks do stall in TR tracts, it may be possible to reverse them without the involvement of RecA or dissociation of the replisome, making resumption of replication possible without primosome re-assembly at RecA-mediated D-loops. In addition, unwinding of the DNA secondary structure by RecQ may be possible, allowing removal of the replication block without cleavage and repair by homologous recombination.

Chapter 6: Competition between plasmids with and without trinucleotide repeats.

Aims

Multimerisation studies performed in chapter 4 of this thesis suggested that DNA ends might be generated at $(CTG)_{43}$ tracts in *E. coli*. If TR tracts can hinder replication forks, are prone to double strand breaks, or induce recombination at DNA secondary structures, they would be expected to be a hindrance to the maintenance of their host DNA molecule. The work in this chapter uses a series of experiments in which competition between plasmids containing TR tracts and pUC18 control is assessed in an attempt to establish whether or not TR tracts are detrimental to the propagation of DNA.

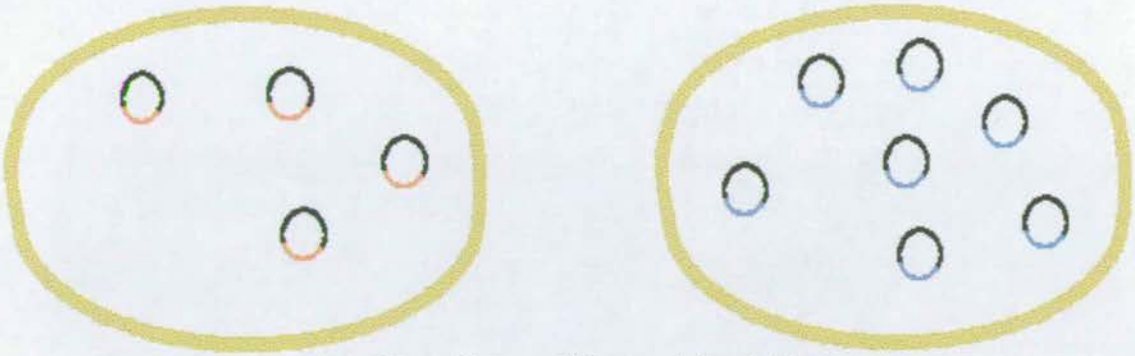
Previous work in this thesis has shown that TR sequences cloned within pUC18 do not induce the SOS response within *E. coli* (see chapter 5), so it was thought that a $(CAG/CTG)_{43}$ tract cloned into a plasmid would not inhibit cell division or have any other detrimental effects on host cell viability. Indeed, during log phase growth from overnight cultures, no differences in growth rates were observed between cells carrying any of the different plasmids.

A reduced TR plasmid yield might be expected if a cell retains the TR plasmid at a much lower copy number than normal because the TR limits plasmid propagation. In previous experiments in wild type strains in this laboratory in which plasmids bearing $(CXG)_n$ TR tracts were extracted from 1.5ml overnight cultures, there was no suggestion of decreased TR plasmid yield compared to pUC18 control.

Since TR plasmids do not induce the SOS response in host cells, display normal exponential phase growth, and have not been observed to reduce plasmid miniprep yield, inter-cell competition between two cells containing different plasmids was

thought to be unlikely to reveal any serious defects in plasmid maintenance. A more stringent test would be required if subtle differences in plasmid maintenance exist.

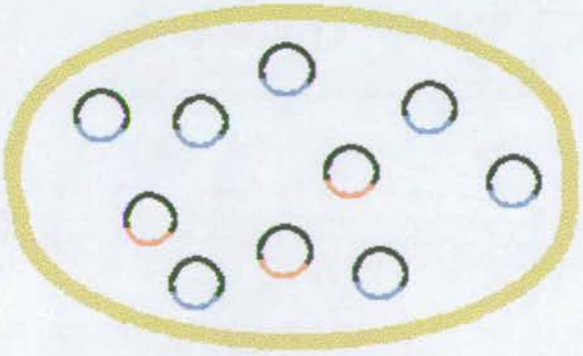
Figure 6.1: Intercell competition.



Intercell competition would probably only result in reduced copy number for a plasmid which is difficult-to-maintain

Competition between different plasmids in the same cell (intra-cell competition) should be more intense as the two types of plasmid would have to compete for the replication and recombination proteins present in each cell. Relative gains or losses of each plasmid type in a cell would be amplified in the culture as a whole. Therefore, the experiments presented in this chapter made use of conditions which enhanced the possibility of intracell competition.

Figure 6.2: Intracell competition.



Intracell competition would be more intense, since both types of plasmid would have to compete for the limited resources within the same cell.

Since antibiotics have different points of action and vary in heat-stability, wild type cells have different tolerances to different antibiotics. The ability to screen for the same selectable marker on both plasmids would eliminate the possibility that one type of plasmid is undergoing more stringent selection than the other. Using the same selectable marker in each plasmid (ampicillin in this case), would also make the experiment more sensitive, as a badly maintained plasmid could be completely removed from each co-transformed cell in the population, rather than simply experiencing a decrease in copy number. By selecting for ampicillin resistance, cells containing just one type of plasmid or any ratio of both plasmids combined could theoretically survive equally well in LB containing ampicillin. With a combined copy number of between 500 and 3000 plasmids per cell, even small differences in plasmid maintenance should be detected. It should be noted that the use of the same selectable marker would allow segregation of the two competing plasmids into separate daughter cells, and may result in inter-cell competition at some stage of the incubation.

It was decided to monitor the difference in yield of the two plasmids co-transformed into the same cell, before and after extended culture. The culture would be deliberately allowed to repeatedly cycle through stationary phase and log phase growth. In this way, the plasmids would be exposed several times to the different types of internal cellular environment found at these culture densities, allowing detection of any difficulty in propagation of TRs through all phases of cell growth.

Competition studies were originally performed with a variety of TR plasmids in wild type AB1157. Subsequent inter-cell studies by Xeufeng Pan in JM83 background confirmed these general trends in many other strains, except *umuDC*⁻ (data not shown). Following the experiments carried out in wild type cells presented here, plasmids exhibiting the clearest competition results were selected for further inter-cell competition experiments in an *umuDC*⁻ background.

Methods

For fair competition between the plasmids, all were prepared in the same *recA*$^-$ strain to give almost exclusively monomeric species. Previous work in chapter 4 established that multimerisation of plasmids bearing CTG_{43} and CAG_{43} repeat arrays was indistinguishable from that of unmodified pUC18 in wild type and *recA*$^-$ cells. No differences in plasmid multimerisation were observed in any of the other plasmids examined in this study. This is of crucial importance to intra-cell competition, because multimers out-replicate monomers and accumulate clonally within the culture (Summers *et al.*, 1993). Paradoxically, cells bearing predominantly plasmid multimers display significantly increased rates of plasmid loss in daughter cells (Summers, 1998), making them more prone to antibiotics when re-inoculated into fresh medium. None of these factors were expected to contribute significantly to any competition observed, thanks to earlier experiments showing no variation in plasmid multimerisation in both wild type cells and the *recA*$^-$ strain used to produce the monomeric plasmids prior to transformation.

Whilst selecting only for ampicillin removes a number of variables affecting competition within the cell, it raises the problem of how to obtain two different plasmids within the same cell when only able to select for one antibiotic resistance. This problem was partially overcome by the manipulation of a number of factors influencing transformation. Electroporation was used to increase the overall transformation frequency, increasing the chance of placing both types of plasmid within the same cell. 10µl of DNA solution containing approximately equal levels of the two plasmids was used to maximise the amount of DNA available for the transformation process without biasing the chance of only one type of plasmid entering a cell. A small volume of electrocompetent cells (20µl) was used to maximise the plasmid to cell transformation ratio. The cell/DNA transformation mixture was allowed at least half an hour incubation on ice before electroporation to maximise plasmid binding to cells. After electroporation, segregation of the two plasmids was reduced by not shaking the recovering cells, so transformed cells were less likely to divide whilst plasmid copy numbers were still low.

Whilst a high transformation frequency is desirable for co-transformation, a dense plating culture would cause difficulties in selecting single colonies. To guarantee good colony separation, the recovering cell culture was sufficiently diluted prior to plating out. Twenty isolated single colonies were chosen for screening and inoculated into 5ml overnight cultures containing ampicillin. After overnight incubation at 37°C with shaking, 1.5ml of culture was extracted to perform plasmid purification using a Qiagen miniprep kit. A PvuII restriction digest was performed to excise a fragment containing the multiple cloning site of pUC18. This fragment was 322 bases in pUC18, and longer for each TR plasmid. The mouse flanking sequences amounted to 76bp. DNA from each plasmid could be separated on a 1.6% agarose gel and stained with ethidium bromide for swift diagnosis of co-transformants. The size of the PvuII fragment for each plasmid used in this study is given in table 6.1.

Table 6.1: PvuII fragment size for each TR plasmid.

plasmid	Total size of PvuII fragment
pUC18	322bp
mouse CAG_{43} and CTG_{43}	527bp
synthetic CCG_{24} and CGG_{24}	394bp
artificial CAG_{28} and CTG_{28}	406bp
mouse CAG_{25} and CTG_{25}	397bp

The time period for competition to occur was 12 days in each experiment, during which time, 50µl of overnight cultures were re-inoculated four times into 5ml of fresh LB with 5µl ampicillin (100mg ml^{-1} stock) to prevent plasmid loss. Following this period of competition, 1.5ml of culture was used to extract a plasmid miniprep, giving an end-sample for the study. To ensure a good recovery of plasmid from the end sample, miniprep samples were always taken from cell cultures that had been re-inoculated into LB-ampicillin the previous day.

Start and end samples for the same culture were run in adjacent lanes on the same 7% agarose gel and stained with Vistra Green for one hour before quantification. The gels were scanned on a Storm PhosphoImager, and band intensities quantified using Imagequant software. Band intensities were normalised for the plasmid-specific size of PvuII cleavage product. The molar fraction of each TR plasmid in the culture before and after the competition period could then be calculated.

Results

The example agarose gel shown below (Figure 6.3) illustrates how beginning and end samples were run in adjacent lanes. There is a greater frequency of CAG_{43} and CTG_{43} plasmid loss compared to CCG_{24} and CGG_{24}, when each competes with pUC18.

Figure 6.8

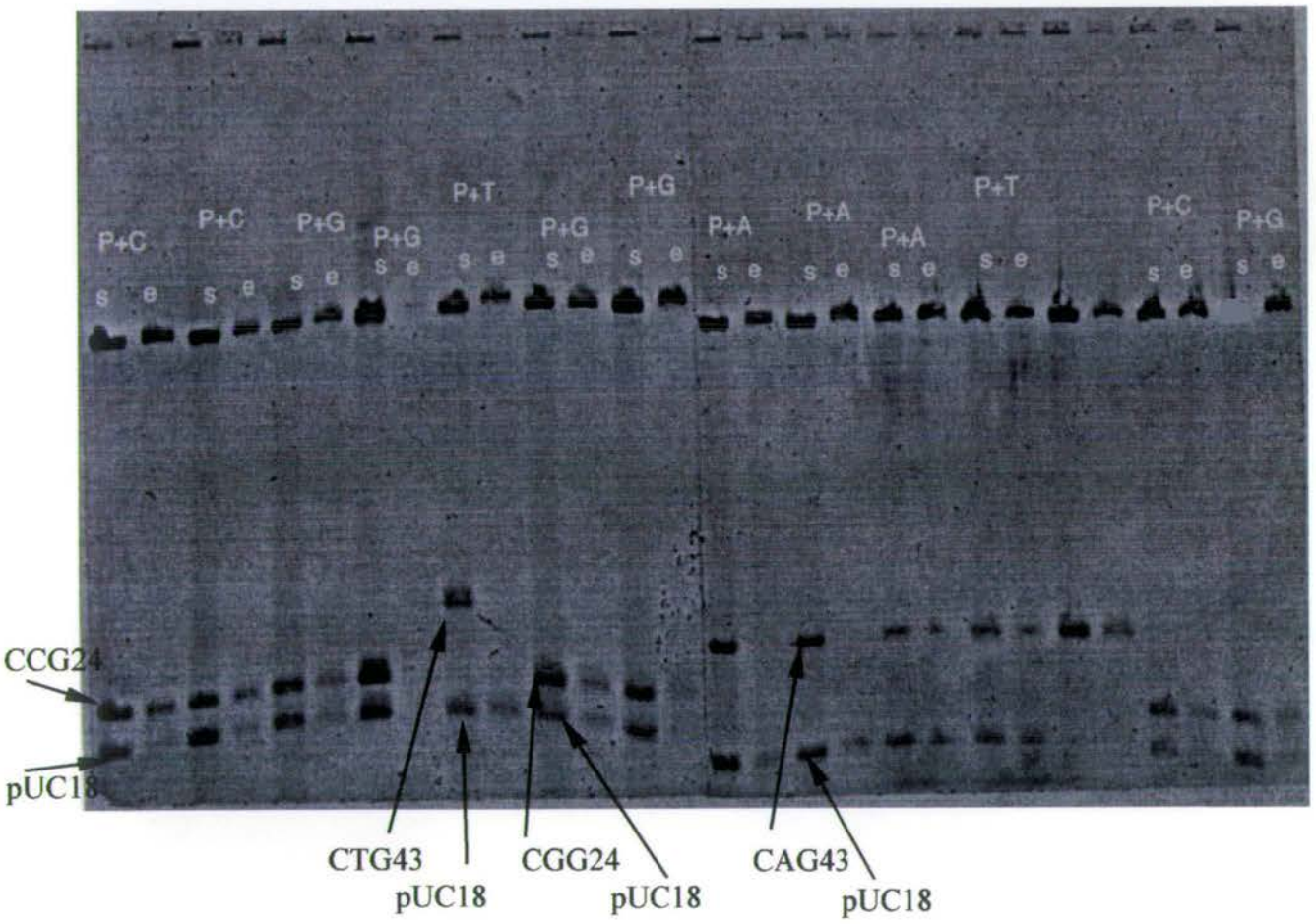


Figure 6.4: The proportion of artificial CCG24 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

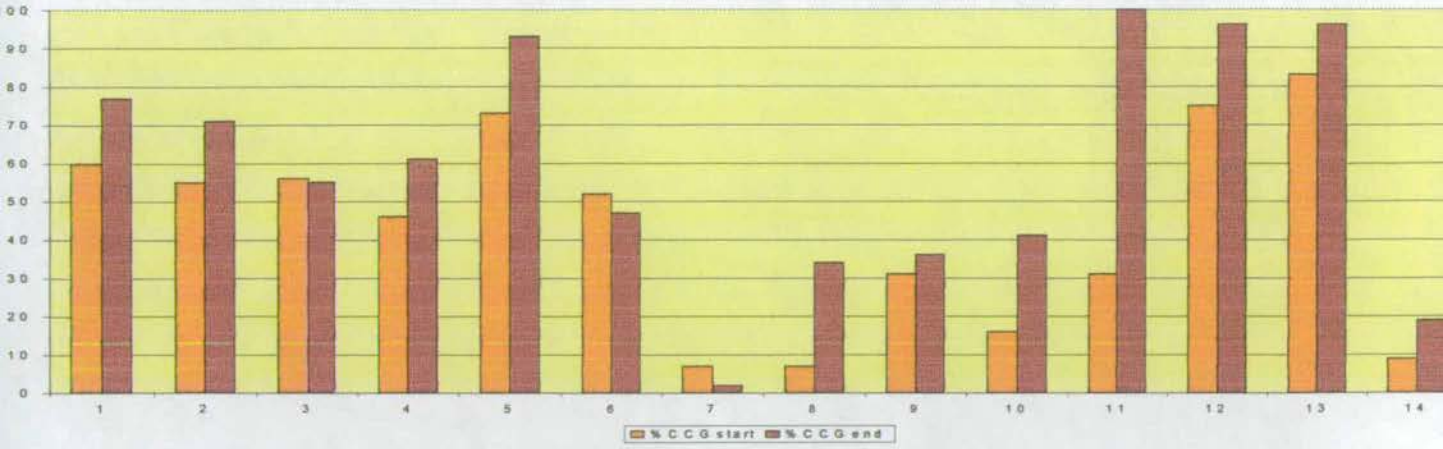


Figure 6.5: The proportion of artificial CGG24 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

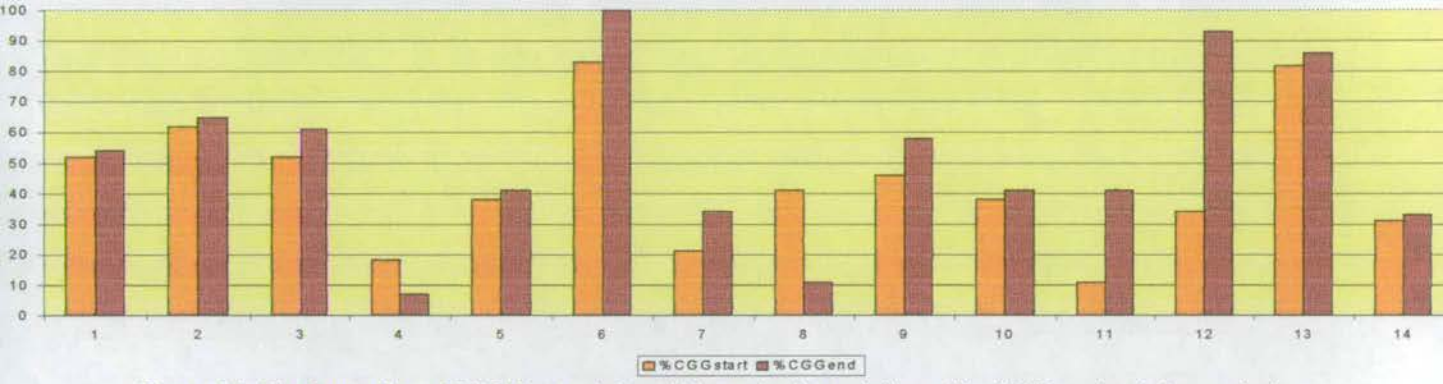


Figure 6.6: The Proportion of CAG43 repeat plasmid in a mixed population with pUC18 vector, before and after a 12 day incubation in wild type cells.

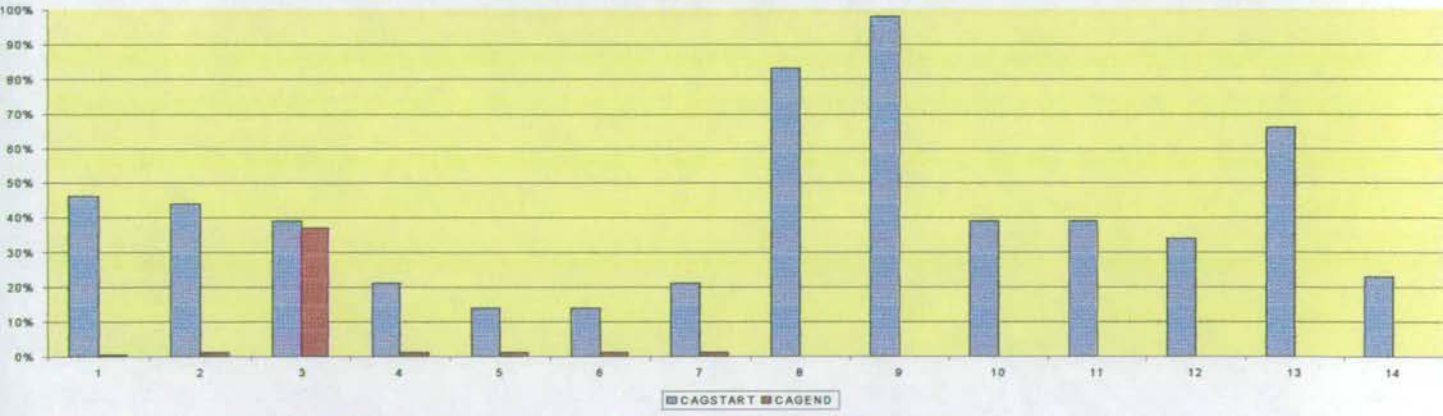


Figure 6.7: The proportion of mouse CTG43 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

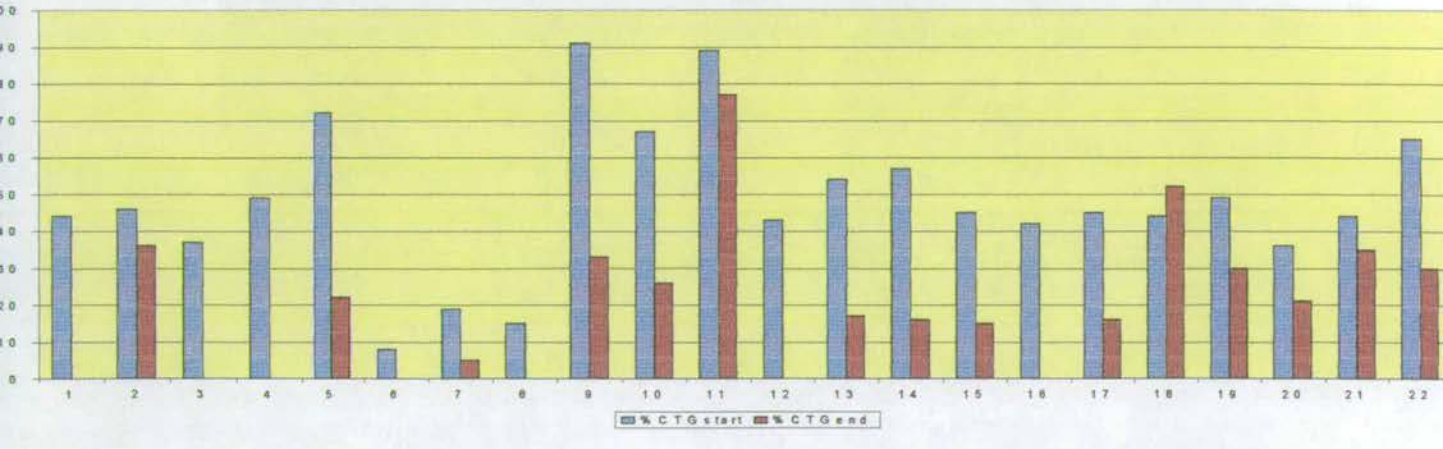


Figure 6.8: The proportion of artificial CAG28 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

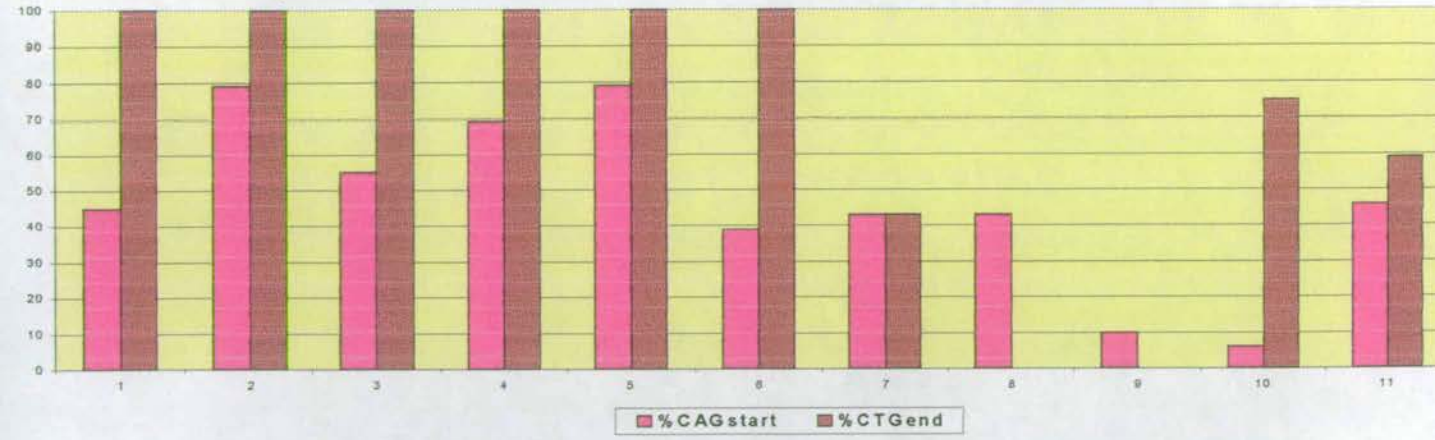


Figure 6.9: The proportion of artificial CTG28 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

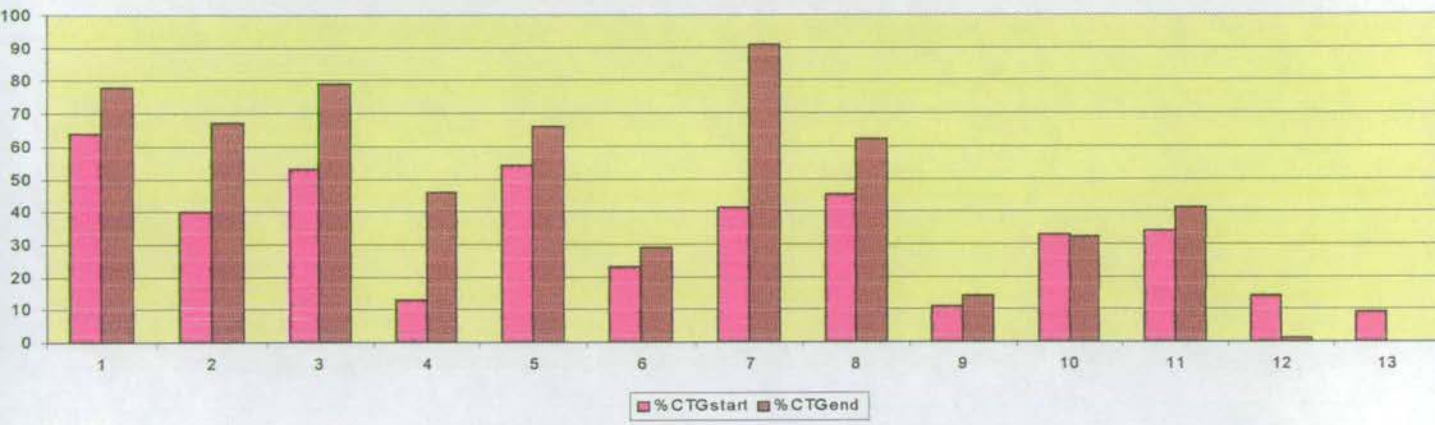


Figure 6.10: The proportion of mouse CAG25 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

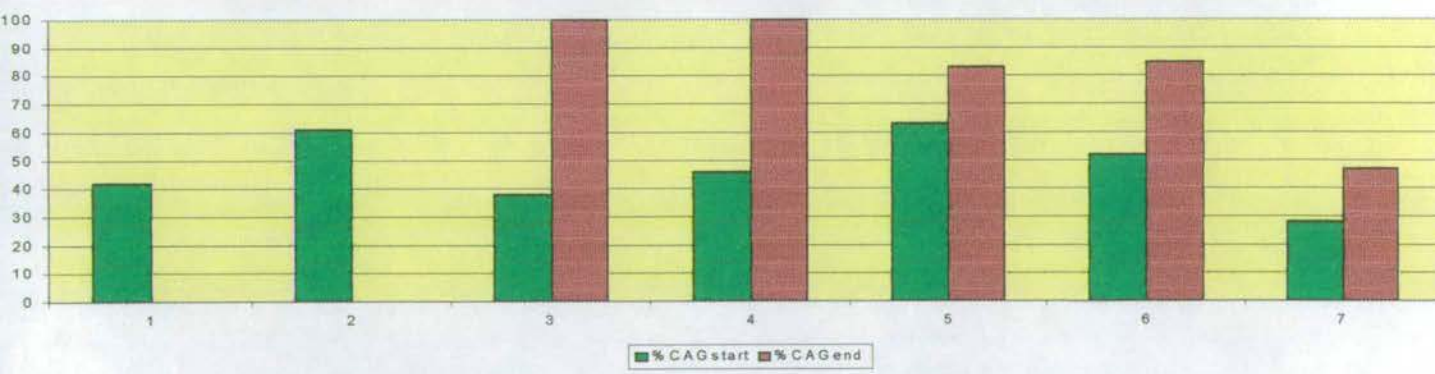


Figure 6.11: The proportion of mouse CTG25 in a mixed population with pUC18, before and after a 12 day incubation in wild type cells.

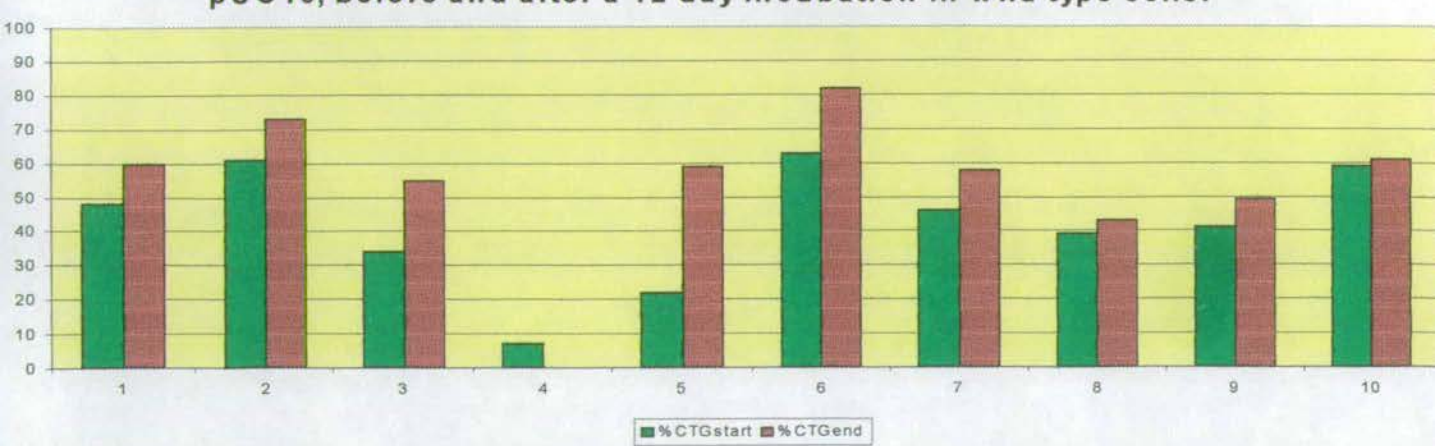


Figure 6.12: The proportion of mouse CAG43 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells.

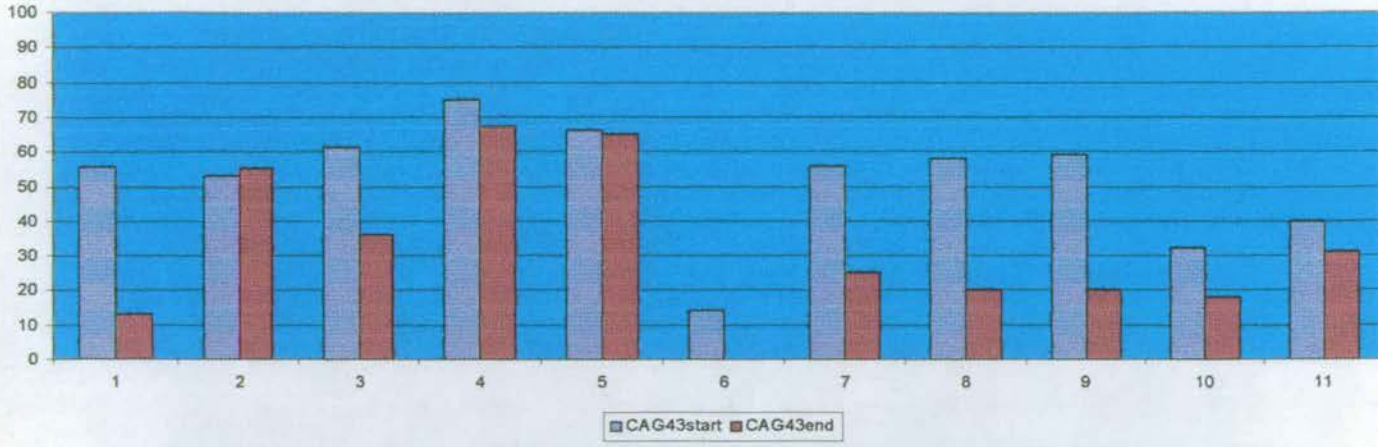


Figure 6.13: The proportion of mouse CTG43 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells.

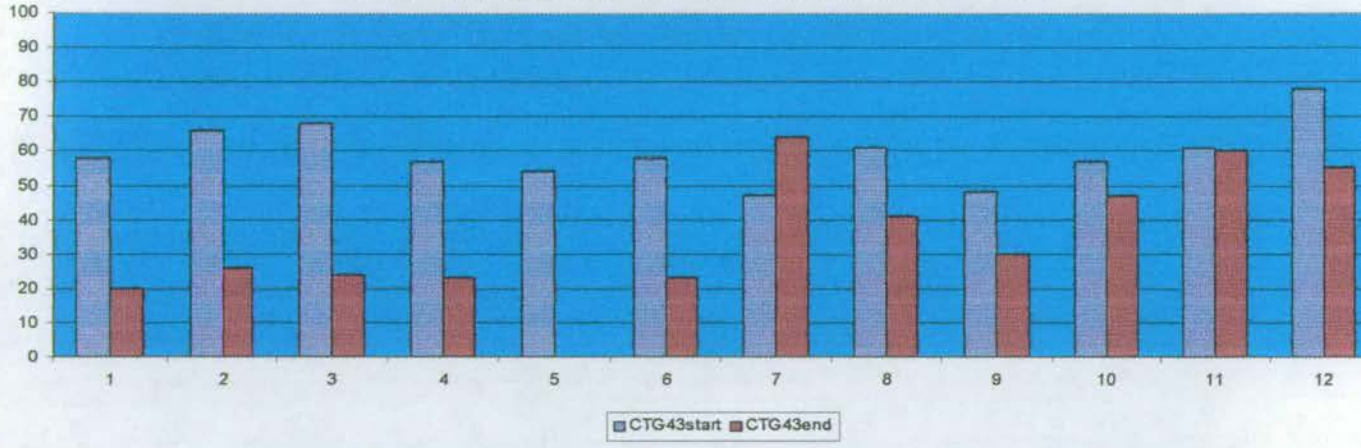


Figure 6.14: The proportion of artificial CCG24 in a mixed population with pUC18, before and after a 12 day incubation in umuDC- cells.

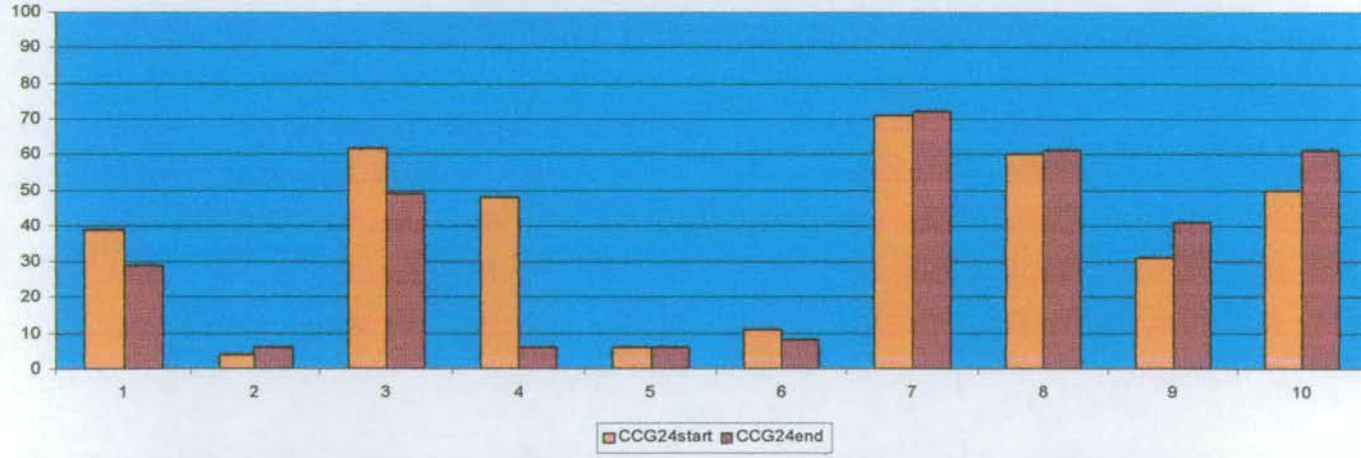


Figure 6.15: The proportion of CGG24 in a mixed population with pUC18, before and after a 12 day incubation in umDC- cells.

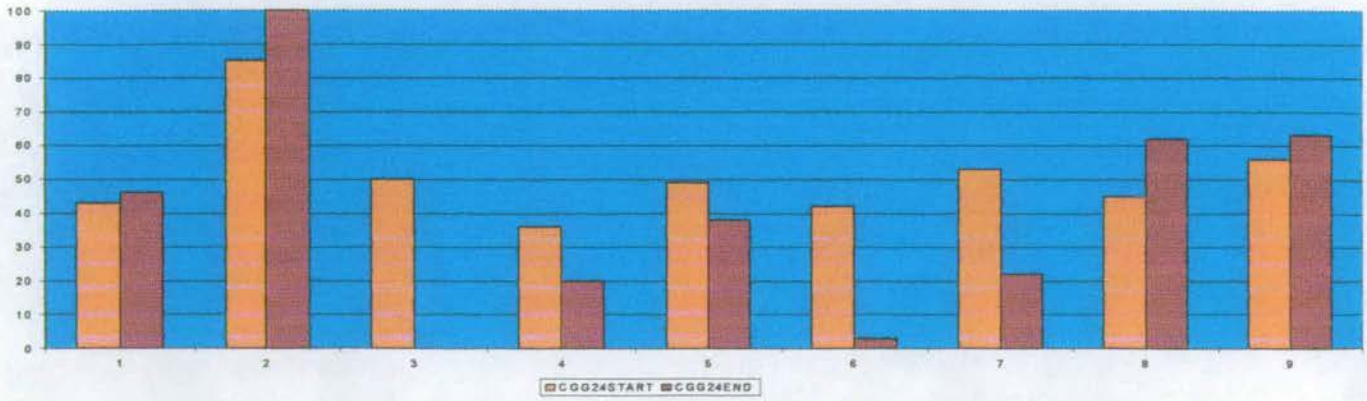


Figure 6.16: mean normalised competition results

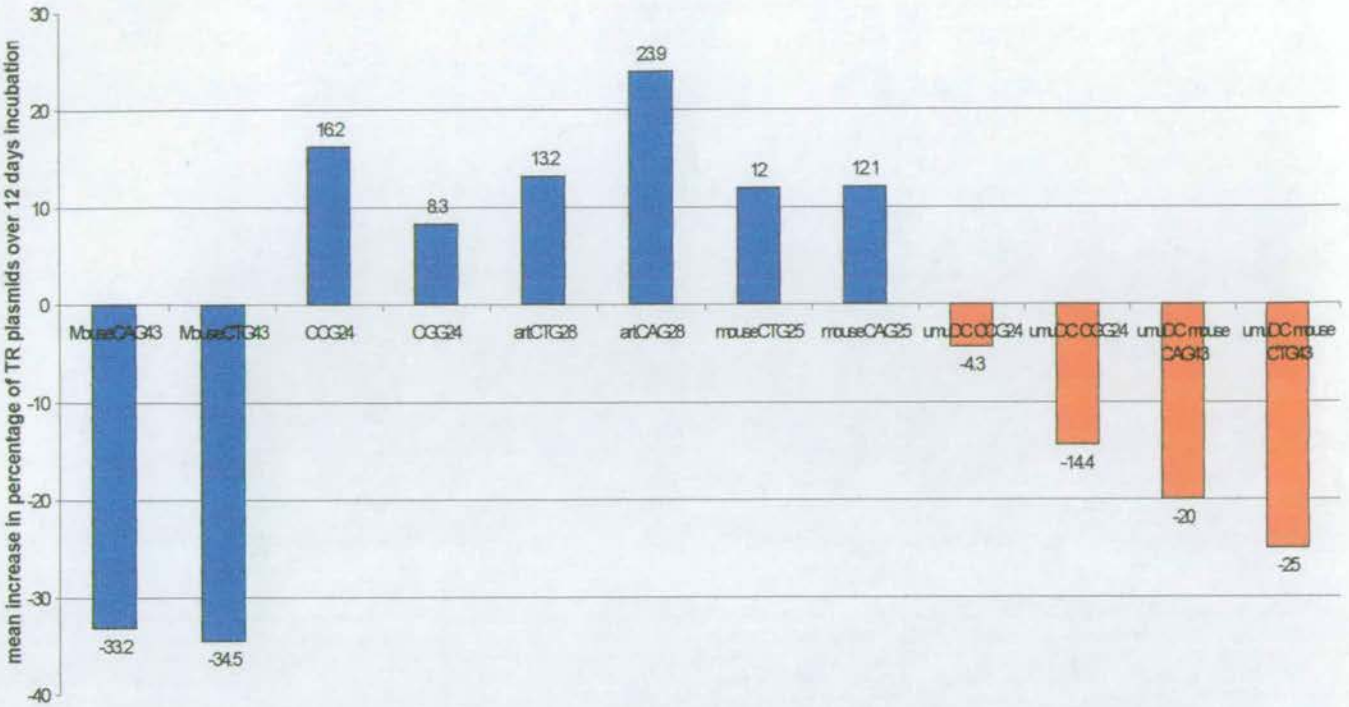


Table 6.2: The mean values of plasmid loss or gain for each plasmid tested

Plasmid	N	Mean	Median	StDev
Artificial CCG_{24}	14	+16.21%	+15.50%	18.38%
Artificial CGG_{24}	14	+8.29%	+3.50%	19.95%
Artificial CAG_{28}	11	+23.9%	+21.0%	33.5%
Artificial CTG_{28}	13	+13.23%	+12.00%	17.54%
Mouse CAG_{25}	7	+12.1%	+20.0%	46.7%
Mouse CTG_{25}	10	+12.00%	+12.00%	12.00%
Mouse CAG_{43}	16	-34.50%	-31.50%	26.81%
Mouse CTG_{43}	22	-28.64%	-32.50%	17.32%
UmuDC⁻ Artificial CCG_{24}	10	-4.30%	+0.50%	15.22%
UmuDC⁻ Artificial CGG_{24}	5	-14.4%	-15.0%	38.9%
UmuDC⁻ Mouse CAG_{43}	11	-20.00%	-14.00%	15.94%
UmuDC⁻ Mouse CTG_{43}	12	-25.00%	-28.50%	20.05%

Sources of variation

The TR plasmids certainly appear to be behaving differently from pUC18 and each other. Whilst trends are visible in the data, it should be noted that the results within some groups are very variable. There are several unavoidable sources of variation within this set of experiments:

1. The fact that uneven numbers of plasmids might be introduced into the same cell (giving one plasmid or another an initial advantage in replication and recombination) might explain some of the variation within the start miniprep for each plasmid-type. Any initial advantage is unlikely to be maintained in the case of troublesome plasmids because of the high copy number of the plasmids and the many generations used in this experiment, so trends in competition should still be apparent. However, it was clear that plasmids with a very low initial representation (usually under 10%) were far more likely to be lost, presumably by genetic drift mechanisms, regardless of how well they competed with pUC18.
2. Occasional colonies show less intense competition than expected (e.g. colony 3 in mouse CAG_{43} wild type). It is possible that the two plasmids were introduced into two separate cells that landed on an agar plate in close proximity, and grew into a colony of cells indistinguishable in shape from one formed from a single founder cell. Alternatively, the two plasmids might have been transformed into opposite ends of a single large cell that was about to divide. Following division, the two daughter cells containing different plasmids would be present at the same location on the agar, restricting the subsequent culture to inter-cell competition. Likewise, inter-cell competition might have occurred at variable times during the incubation due to random segregation of the two plasmid types into separate daughter cells.
3. It is important to note that competition actually started immediately after both plasmids were introduced into the same cell. Therefore competition had already occurred during the generations required to form a colony on LB agar and subsequent overnight incubation in LB before the "start" miniprep sample could be taken. This is one reason why the initial plasmid representation was never 50% for every TR tested. It is interesting to note that during this initial period of competition before the first miniprep sample was taken, the culture would not have entered stationary phase.

The sign test

These high levels of variation within the groups make statistical analysis of the raw data problematic. Although not very powerful, the sign test was considered the most appropriate test here to aid statistical analysis. In this test the number of replicates for which there is an increased percentage, decreased percentage, or no change, are counted and tested using a binomial distribution with p=0.5. If a TR plasmid is just as easy to maintain as pUC18, then just by chance approximately half the observed differences will be positive and half will be negative. For example, mouse CTG_{25} is included as significant in this test because it has consistent but small increases in culture compared to pUC18.

Table 6.3: Sign test values for TR plasmid normalised competition results.

	N	Above	Equal	Below	P
Artificial CCG_{24}	14	11	0	3	0.0574
Artificial CGG_{24}	14	12	0	2	0.0129
Artificial CAG_{28}	11	8	1	2	0.1094
Artificial CTG_{28}	13	10	0	3	0.0923
Mouse CAG_{25}	7	5	0	2	0.4531
Mouse CTG_{25}	10	9	0	1	0.0215
Mouse CAG_{43}	16	0	0	16	0.0000
Mouse CTG_{43}	22	1	0	21	0.0000
UmuDC Artificial CCG_{24}	10	5	1	4	1.0000
UmuDC Artificial$(CGG)_{24}$	5	1	0	4	0.3750
UmuDC Mouse CAG_{43}	11	1	0	10	0.0117
UmuDC Mouse $(CTG)_{43}$	12	1	0	11	0.0063

The sign test indicates that mouse CTG_{25}, artificial CGG_{24}, and possibly artificial CCG_{24}, have significantly increased their representation within the culture by the end of the 12 day incubation. In contrast, mouse CAG_{43}, and mouse CTG_{43}, both display decreased percentages in the final samples from both wild type and *umuDC*$^-$ cells, though the decrease in *umuDC*$^-$ cells is not as large as in wild type.

Total plasmid size was not a factor in determining the outcome of competition.

It is possible that the difference in plasmid yields over the course of this experiment could not only be due to the presence or absence of the TR, but also in the size of the plasmids replicated. The competition results for plasmid CAG_{43} with mouse flanking sequences is particularly striking. Out of 16 wild type co-transformants, 15 show complete loss of mouse CAG_{43}. This plasmid is 205bp larger than the 2686bp of unmodified pUC18, and it seems unlikely that an increase in size of just 7% should have such a dramatic effect on its ability to compete with pUC18, even over the many bacterial generations available for plasmid selection in this study. This belief is reinforced by competition results from artificial CCG_{24}, artificial CGG_{24}, artificial CAG_{28}, artificial CTG_{28}, mouse CAG_{25}, and mouse CTG_{25}, all of which are larger than pUC18, yet display varying tendencies to out-compete pUC18 or at least maintain their original representation within wild type cells. Therefore it would seem that the loss of mouse CAG/CTG_{43} TR plasmids is a genuine effect independent of plasmid size.

Loss of mouse CAG/CTG_{43} did not result from deletion of the TR tract.

In mouse CAG_{43}, the complete absence in the end minipreps of PvuII fragments other than that expected from pUC18, was informative. It indicated that this plasmid had not simply undergone recombination mediated deletion events to remove the TR tract (making the pUC18 band effectively a mixture of pUC18 and deleted $(CAG)_{43}$-bearing plasmid). Complete deletions would leave the mouse flanking sequences within the multiple cloning site. No smears or additional bands were present in the gels for any of the samples analysed from any of the plasmids, but confirmation of this lack of deletion would require a more sensitive assay such as the end labelling carried out in chapter 7.

The effect of triplet sequence on competition.

Both artificial CCG/CGG_{24} and mouse CAG/CTG_{25} display modest gains over pUC18. Their overall behaviour appears to be rather similar. However, this can only be said for plasmids bearing these relatively short repeats. More pronounced differences might be evident for longer repeat tracts with greater propensity for more stable folding. As such, these results do not provide any insights into different forms of secondary structure or propensities for formation characteristic to a particular TR sequence.

TR orientation has little effect on competition.

The mean competition results appear to indicate that orientation has little effect on competition in mouse CAG/CTG_{25} and mouse CAG/CTG_{43}. In the case of the longest TR tracts, the sign test results for the mouse CAG_{43} repeats are almost identical to mouse CTG_{43}. Orientation dependence is more evident for the mean competition results in artificial CCG/CGG_{24} and artificial CAG/CTG_{28}, though the two orientations of these TRs do display similar general trends of loss and gain, and the sign test results for artificial CAG/CTG_{28} plasmids are similar. In general, the mean competition results appear to indicate that the orientation of the TR is not as significant as TR length in determining the outcome of competition. Consequently, competition results can be assessed independent of TR orientation to increase the sample size of these studies, and further clarify the observations made.

Table 6.4: Outcomes of TR plasmid competition independent of TR orientation.

	N	Above	Equal	Below	Mean gain	Median gain	P-value
Artificial CCG/CGG_{24}	28	23	0	5	13%	12%	0.0004
Artificial CAG/CTG_{28}	24	18	1	5	12%	13%	0.0080
Mouse CAG/CTG_{25}	17	14	0	3	11%	12%	0.0052
Mouse CAG/CTG_{43}	38	1	0	37	-35%	-37%	0.0000000001
UmuDC Artificial CCG/CGG_{24}	19	9	1	9	-12%	-11%	0.1762
UmuDC Mouse CAG/CTG_{43}	23	2	0	21	-23%	-23%	0.00003

The presence of mouse flanking sequences is not an important factor in competition.

In wild type cells, the median gain of artificial CAG_{28} is +21%, which is very similar to the mouse CAG_{25} median value of +20%. Similarly, the wild type median value for artificial CTG_{28} is +12%, which is highly comparable to the mouse CTG_{25} median value of +12%. Thus the mouse flanking sequences can be said to have no effect on plasmid maintenance in this assay, and certainly less of an effect than trinucleotide repeat length. (See appendix 1 for the mouse flanking sequences).

Reduced competition was observed in the *umuDC*⁻ background.

The relative loss of CCG/CGG_{24} in *umuDC*⁻ cells suggest that the relative gain of artificial CCG/CGG_{24} plasmids in wild type cells is *umuDC*⁺-dependent. However, the sample size of CCG/CGG_{24} in *umuDC*⁻ cells is rather small. In contrast, the larger sample size of mouse CAG/CTG_{43} in *umuDC*⁻ cells give sign test results indicating that mouse CAG/CTG_{43} continues to lose to pUC18 in an *umuDC*⁻ strain, but to a lesser degree than in wild type cells. It would appear that in wild type cells the long TR tract has an *umuDC*⁺-dependent maintenance disadvantage compared to pUC18. Unfortunately the relative change in competition is impossible to determine accurately from this experiment. Complete loss of mouse CAG/CTG_{43} may have occurred very early in the 12 day incubation in wild type cells, but was not detected until the end samples were taken. A time course experiment involving samples taken at shorter intervals (for example, the day after each re-inoculation) would be required to provide this information.

Discussion

The relative loss of mouse CAG/CTG$_{43}$

The sign test values for competition in wild type cells indicate that the mean loss of mouse CAG/CTG$_{43}$ plasmids is highly significant. Clearly CAG$_{43}$ and CTG$_{43}$ repeat tracts with mouse flanking sequences strongly inhibit the maintenance of their host plasmids in *E. coli*. Since this effect is not observed in mouse (CAG/CTG)$_{25}$ repeats, it was interpreted as a length-dependent effect of the TR tracts. Additionally, identical experiments performed with plasmids containing less than thirty consecutive triplets (CAG/CTG$_{28}$, and CCG/CGG$_{24}$) could not duplicate the TR plasmid loss observed with mouse CAG/CTG$_{43}$.

Possible loss via inhibited replication through long TR tracts

If hairpin structures formed in TR tracts do stall replication forks, replication of TR plasmids would be slower, and may have increased rates of collapse. Results from the SOS reporter assay experiments performed in chapter 5 indicate that if replication disruption does occur in (CAG/CTG)$_{43}$ tracts, the resulting DNA structures are not persistent substrates for RecA*. Polymerase uncoupling would result in single strand gaps, and replication fork collapse would result in DSBs. Therefore, if replication fork stalling does occur in these TR tracts, re-initiation without the need for recombination is more likely. When considering this, the mechanism of replication inhibition in the longest TR tracts examined here is more likely to be fork pausing rather than collapse. In support of this possible explanation of CAG/CTG$_{43}$ loss, inhibition of DNA synthesis has previously been observed in TR tracts *in vitro* (Ohshima and Wells, 1997 and 1998; Kang *et al.*, 1995) and inhibition of DNA replication *in vivo* (Samadashwily *et al.*, 1997).

In addition to direct replication blockage by secondary structures formed in TR tracts, there is evidence that DNA hairpins pause RNA polymerases and terminate transcription (reviewed in Platt, 1986). This may result in stable complexes which in turn hinder the progress of replication attempting to pass through them. Additionally, transcription through a simple mono-nucleotide repeat d(C)$_{32}$ transcription template

has been observed to stall transcription, resulting in an RNA/DNA complex (possibly including the RNA polymerase), which stalls DNA replication (Krasilnikova *et al.*, 1998). The replication block was found to be repeat length and orientation dependent, but not dependent on orientation relative to the origin of replication. The exact nature of this replication block is unknown, but suggested structures include parallel triplex and collapsed R-loops. The combination of transcription inhibition at DNA hairpins and at simple direct repeats makes it likely that TR tracts also inhibit transcription. Indeed, several studies have linked transcription to elevated deletion frequencies in $(CTG/CAG)_n$ tracts (Schumacher, Pinet, and Bichara, 2001; Bowater *et al.*, 1997), an observation which may be explained by elevated rates of replication fork pausing.

Possible loss via endonucleolytic cleavage of hairpin structures in TR tracts

Alternatively, it has been proposed that palindromes and by analogy TR tracts do not inhibit replication directly by blocking the passage of replication forks, but form hairpin structures which act as substrates for cleavage and subsequent recombination (Cromie *et al.*, 2000). It is possible that the loss of mouse $(CAG/CTG)_{43}$ is related to the production of DNA ends thought to occur in these plasmids (revealed by work in chapter 4 of this thesis). DSBR by RecBCD would probably not be successful in the absence of *chi*, but might simply degrade the plasmid. In this way, a propensity for DSBs in long TR tracts would make the TR plasmids more susceptible to degradation by the ds exonuclease activity of RecBCD, and give pUC18 a competitive advantage.

By analogy with hairpin processing in inverted repeat sequences, SbcCD would be the most likely cause of DSBs generated by nuclease action in TR tracts. Reduced loss of CAG/CTG_{43} plasmids in competition experiments carried out in an *sbcCD*$^-$ background would answer this specific question. However, other activities such as MMR or ss endonucleases might be able to produce off-set nicks in TR tracts which may be converted into DSBs if not ligated before replication.

Possible loss via the formation of linear multimer plasmids

It is also possible that TR plasmids are lost from this experiment in the form of linear mulimers. TR plasmids attempting recombinational repair using an intact copy of itself or pUC18 as a template might initiate rolling circle replication. This would not induce the SOS response, but would require RecA. A double rolling circle has two protected ends, so is resistant to degradation by RecBCD. Daughter cells receiving only long multimers would be prone to plasmid loss, allowing preferential removal of TR plasmids from the population of cells.

ColE1-type plasmids (such as pBR322 and pACYC184) do not normally undergo rolling circle replication, but have been observed to do so in cells expressing inhibitors of RecBCD (Bravo and Alonso, 1990), and in *recBC*$^-$ cells where approximately 14% of all the plasmids are linear multimers, and in *recBC*$^-$*sbcB*$^-$ strains where they account for 74% of all the plasmids present (Cohen and Clark, 1986). RecBCD may be inhibited in SOS-induced cells (see iSDR relevance to short TR plasmid gain, later). Rolling circle replication has been demonstrated to be partially insensitive to chloramphenicol (Leonhardt *et al.*,1991), raising the possibility that it can occur under conditions of limited protein synthesis. The combination of these findings make it tempting to speculate that the formation of linear multimers of plasmids capable of secondary structure would be most likely to accumulate in stationary phase cells known to be SOS-induced (Taddei, Matic, and Radman, 1995). However, for this to be an adequate explanation of CAG/CTG$_{43}$ loss, it must account for more frequent rolling circle replication initiation by long TR tracts.

The method of replication initiation in some specialised plasmids using rolling circle replication could act as a model for the possibility that hairpin processing leads to some rolling circle replication in plasmids. Studies of pT181 describe an interesting form of replication initiation in which a protein dimer called RepC/C specifically extrudes and stabilises a GC-rich DNA cruciform (Wang *et al.*, 1993). This enables specific nicking of the ss loop structure at the top of the hairpin, allowing one end of the nick to re-bind a template strand and be used as a primer for DNA Polymerase III (Jin *et al.*, 1997). Cruciform extrusion also melts an area of DNA necessary for replisome assembly (Jin *et al.*, 2001). Whilst this is clearly a specialised form of

plasmid replication, it does illustrate that hairpin processing can lead to break-dependent replication initiation common to all rolling circle replication (Viret, Bravo and Alonso, 1991). Indeed, mapping the start points of plasmid rolling circle replication in *B. subtilis* revealed that 25% of the DNA ends originated in sigma origins of replication, but were independent of orientation and therefore sigma-replication activity (Leonhardt *et al.,* 1991). It was assumed that the sigma origin of replication present in either orientation was able to form a distinct DNA structure that was susceptible to nicks or DSBs, and acted as a recombination hot-spot from which rolling-circle replication could be initiated.

Possible loss of long TR plasmids by delayed recovery from stationary phase

An alternative mechanism of selection against long TR plasmids was suggested by Bowater and co-workers (1996). It was proposed that CTG repeats in plasmids were observed to affect the entry of *E. coli* into exponential growth phase by increasing the lag phase in a manner proportional to TR tract length. That study confirmed that the length of the CTG repeat had no effect on the rate of exponential growth or cell viability (the only exception being plasmids bearing very long TR tracts of $(CTG)_{175}$, which displayed approximately a tenfold reduction in the viability of cells in stationary phase). Inter-cell competition results from the same study showed that cells with plasmids bearing longer CTG repeats (175 versus 17 triplet repeats) were propagated more slowly overall, but only if cells passed through stationary phase. The mechanism of extending lag phase was unknown at the time of the Bowater study (1996), but more recent work has found that replication can undergo an UmuD-mediated "timed pause" after DNA damage (see later).

Cells that contain just pUC18 (through random segregation) could have a shorter lag phase upon culture re-inoculation into fresh LB. The growth advantage of cells containing just pUC18 would be exponentially proportional to the lag phase extension and determined by CTG repeat length. This could explain why pUC18 has a replicative advantage compared to mouse CTG_{43} and not mouse CTG_{25}. Unfortunately, it is unknown what proportion of the final cell culture contained each plasmid, and what proportion still contained both plasmids. Consequently, the extent

of plasmid segregation is unknown. However, this "delayed shift" model for the relative loss of CAG/CTG_{43} does at least appear to be specific for long TR tracts, so further investigation was undertaken to test it in an *umuDC*⁻ strain (see later).

The relative gain of short TRs.

The gain of artificial CCG/CGG_{24} plasmids in wild type cells is a dramatic and unexpected result. It is implausible that pUC18 has elevated rates of DNA damage compared to identical but larger plasmids containing TRs, so this effect is likely to be a result of short TR plasmids possessing elevated rates of replication. Since the presence of short TR tracts is unlikely to enhance the processivity of existing replication forks, replicative advantage is most likely to be achieved by TR plasmids initiating replication at elevated frequencies. Plasmids with and without TRs extracted from wild type and *recA*⁻ cells demonstrated rates of dimerisation independent of the TR, so all of the plasmids contained a single copy of the same plasmid origin of replication (*rep* from pMB1) responsible for the high copy number of pUC18 (see DNA replication in pUC18-derived plasmids, chapter 1). The TR plasmids and pUC18 are all subject to this same copy number control, so replication priming must occur within the short TR tracts themselves. This could occur from R-loops (as in cSDR), or from D-loops (as in RDR and iSDR), described in chapter 1.

Replication can be initiated from stable R-loops

One form of replication initiation, called constitutive stable DNA replication (cSDR), has been detected at stable R-loops in *recG*⁻ and *rnaH*⁻ mutants. Since the TR DNA is present in the *lacZ* polylinker, it is likely to be transcribed, and it is possible that replication priming occurs from stable transcript R-loops formed within TR tracts. R-loops should be unwound by RecG or degraded by RnaseH1 in a wild type background, but it is possible that these activities could be titred by the high copy number of pUC18-based plasmids. However the ColE1 origin of plasmid replication is also regulated by an R-loop, yet the copy number of these plasmids are elevated in a *recG*⁻ background, suggesting that in wild type cells, RecG is effective in

suppressing replication initiation even for high copy number plasmids (Vincent, Mahdi and Lloyd, 1996).

DSBR of chi-less plasmids in SOS-induced cells may favour replication (iSDR) rather than plasmid degradation

It is also possible that recombination initiated at some form of substrate within the TR tract generates D-loops which can be used to prime replication. Replication primed at D-loops formed as a result of ss gap repair is restricted to the length of the gap. Double strand ends (DSEs) have been implied at stalled replication forks (Freudenreich 1998, Tishkoff 1997, Michel 1997). However, a DSE induced by replication would not add to a plasmid's competitiveness because replication of the plasmid would be slower overall, even if re-initiation were achieved. This has been demonstrated in the rich-media sensitivity of mutants reducing the processivity of DNA Polymerase III, such as *holD*˙ (Flores *et al.*, 2001). The only way a stalled or broken fork could enhance replication would be if it were to be transformed into two forks travelling in separate directions, but this is not readily explicable. Similarly, an ends-out recombination reaction would establish two replication forks and give two plasmids, but requires one damaged fragment and one intact template plasmid to begin with, so gives no net gain.

However, DSBR is known to initiate longer replication events, and thus could contribute significantly to the replication of 2.7kbp plasmids. Some of the recombination-dependent replication events studied by Kogoma and co-workers were initiated at specific origins as a result of SOS-regulated site-specific endonucleases (Kogoma, 1997). It seems equally possible that nuclease action at secondary structures formed in TR tracts could also initiate replication, thus escaping copy number control determined by regulatory evens at the plasmid origin of replication. However, for this recombination-dependent replication to be possible in the pUC18-derived plasmids used in the present study, the ds exonuclease activity of RecBCD must be reduced.

TR plasmids might suffer DSBs independent of replication, generated as a result of ss endonuclease action at mismatch loops or ssDNA within secondary structures, or by RuvABC cleavage of cruciforms or SbcCD cleavage of staggered hairpins. Perhaps replicative repair of DSBs generated within the short TR tract would be able to initiate two replication forks travelling in opposite directions. However, DSBs in *chi*-less plasmids would normally lead to plasmid degradation by RecBCD, not D-loop formation.

The exception to this DSB-initiated plasmid degradation is in SOS induced cells, where the ds exonuclease activity of RecBCD is inhibited, resulting in inducible stable DNA replication (iSDR) from D-loops. Significantly, Thoms and Wackernagel (1998) observed reduced RecBCD activity (measured by the efficiency of plating of *chi*-less phage) in SOS induced cells. Prior induction of the SOS response was found to increase the efficiency of plating of phage T4 (defective in the protective DNA end-binding protein encoded by *gene 2*). Cellular RecBCD activity is titrated by excision repair-dependent fragmentation of the host chromosome after a massive UV dose, but this effect is enhanced by an SOS-regulated gene product, subsequently identified as SSB. Over expression of SSB from a plasmid was also found to have a similar effect in enhancing T4 growth. SSB had previously been shown to inhibit RecBCD activity at ss overhangs of less than 25 nucleotides (MacCay and Linn, 1976). Undoubtedly, this situation is quite uncommon regarding the generation of so many free ends to titre RecBCD activity, but a similar potential for end-generation might be present in plasmids achieving a very high copy number in stationary phase cells.

Origin-independent plasmid replication has been induced by the artificial generation of DSBs (Asia, Bates and Kogoma, 1994). The DSB is believed to be created by an unknown SOS-induced *oriM*-specific endonuclease. Extensive degradation by RecBCD does not occur due to suppression of its exonuclease activity in SOS-induced cells. This suppression of RecBCD exonuclease activity prevents enlargement of the DSB into a double strand gap, which in standard DSBR restricts polymerase activity to a gap-filling role.

The SOS response is induced in resting bacterial populations (Taddei, Matic, and Radman, 1995). It was observed in ageing colonies after 2 days of incubation on agar plates. The time between re-inoculations into fresh media was 3 days in the competition experiments carried out here, so the SOS response probably would have been induced in these cells also. The cAMP dependence of this SOS induction in ageing colonies suggests that it is mediated by the phenomenon of enhanced RecA expression due to depletion of cellular levels of cAMP in starving cells (Janion *et al.*, 2002). Enhanced RecA expression was examined using an immunoassay, and found to occur after the end of exponential growth, when cells enter the stationary phase. An enhanced rate of generation of rifampicin resistant colonies (typical of mutagenic response) also coincided with this time period, further suggesting SOS induction at this stage (Taddei, Matic, and Radman, 1995). Since SSB is under LexA regulation, it is therefore quite possible that RecBCD-mediated DSBR is repressed in stationary phase cells, allowing iSDR in *chi*-less plasmids.

DSB formation was proposed in chapter 4 in cell cultures that had not passed through stationary phase. DSBR has previously been suggested as a mechanism for gene conversion to explain how large expansions previously observed in TR tracts were no longer seen in *recA*$^-$ or *recBC*$^-$ strains (Jakupciak and Wells, 2000). Neither of the plasmids used in that study contained *chi* sites. However, if iSDR does allow replication initiation from persistent D-loops, it is not clear why this might occur more often in short TR tracts compared to long TR tracts.

Reduced competition in *umuDC*⁻ cells

Competition studies using CAG/CTG$_{43}$ plasmids were carried out in *umuDC*⁻ cells to determine whether the mean loss to pUC18 they exhibited in wild type cells is UmuD-dependent. This is an important question mechanistically, because previous studies (Bowater *et al.*, 1996) have shown that cells carrying long TR tracts in plasmids display extended lag phases when passing from stationary to exponential phase. More recently, work performed in the Walker laboratory suggested that UmuD acts as a prokaryotic DNA damage checkpoint, delaying resumption of DNA replication (Opperman *et al.*, 1999). It is therefore possible that the longer TR tracts activate this DNA checkpoint in wild type cells and thus confer a host cell disadvantage compared to daughter cells containing only pUC18.

During the course of these experiments, the mean loss of CAG/CTG$_{43}$ plasmids was not as severe in *umuDC*⁻ cells as observed in wild type cells. To confirm that this was not a result of elevated multimerisation rates in this background, an identical experiment to those in chapter 4 was performed. The mean multimerisation SCD values were 11% for mouse CAG$_{43}$, 12% for mouse CTG$_{43}$, and 8% for pUC18 control. All of these values remained within error bars for wild type multimerisation values for each plasmid, and as such indicate very similar levels of plasmid dimerisation in *umuDC*⁻ cells, regardless of TR tract. The TR plasmids may have a slight replicative advantage over pUC18 in *umuDC*⁻ cells, but it is not clear whether this factor alone can explain the reduced mean loss of mouse (CAG/CTG)$_{43}$ in *umuDC*⁻ cells compared to in wild type cells (compare graphs 5.1 and 5.2 with 5.9 and 5.10).

During the early stages of SOS de-repression, UmuD accumulates in the cell, and inhibits replication. This results in a "timed pause", giving opportunity for faithful repair of DNA lesions. The duration of this delay was observed to be approximately 25 min for a UV dose of 20J/m^2 and 40 min for 50J/m^2 (Opperman *et al.*, 1999). UmuD-dependent replication pausing was detected by measuring replication rates based on ^{3}H thymine incorporation in to the chromosome, which could be correlated to cell density based on OD600 measurements (Opperman *et al.*, 1999). It is thought that this UmuD-dependent timed pause of replication cannot be removed early by any

particular condition being satisfied by DNA repair. Following accumulation early in the SOS response, UmuD can be removed in one of two ways. If the elevated recombination capacity of the early SOS-induced cell has successfully removed the DNA lesions during the timed pause, UmuD is depleted as LexA-inhibited transcription is reduced, and Lon gradually cleaves existing UmuD. Alternatively, if the DNA insult remains, UmuD is converted into UmuD' to actively prime replication re-start at the expense of point mutations generated during translesion synthesis.

The method UmuD uses to achieve replication pause may be linked to the high affinity that UmuD has for the β clamp processivity subunit of DNA polymerase III. This interaction has been studied in crosslinking experiments, and is mediated by residues 9-39 on the N terminal region of UmuD (Sutton *et al.*, 2001). This is exactly the region lost during UmuD auto-cleavage, and UmuD' has far reduced affinity for β clamps. It seems probable that the loading of fresh β clamps in the region of a disrupted replication fork may be inhibited by UmuD competitively binding DnaN monomers, drastically restricting Pol III processivity to only a few nucleotides. Recent work in the Michel laboratory on RecA-independent replication fork regression has determined a necessity for the β-clamp (Grompone *et al.*, 2002). The β clamp is thought to target a fork reversal helicase to sites of replicative difficulty, and remains bound to the DNA to aid subsequent Pol III re-loading after removal of the DNA lesion.

Although UmuD does appear to have a role in the loss of mouse CAG/CTG_{43} plasmids, an UmuD-dependent extended lag phase in cells containing this plasmid cannot completely account for the loss of CAG/CTG_{43} repeats because they still lose out to pUC18 in *umuDC*⁻ cells.

The absence of UmuDC also has an effect on the shorter TR tract tested. Despite a relatively small sample size, control competition studies of CCG/CGG_{24} in *umuDC*⁻ cells appeared to give mean losses compared to pUC18. This is in sharp contrast to significant gains observed by these plasmids in wild type cells.

SCD dimerisation values in *umuDC*⁻ cells were 7.8% for pUC18, 10.4% for artificial CCG_{24}, and 11.8%for CGG_{24}. The general loss of artificial CCG/CGG_{24} observed in

the *umuDC*⁻ background (compare graphs 5.3 and 5.4 with 5.11 and 5.12) is therefore contrary to their elevated dimerisation rates compared to pUC18 in this background. Consequently, biased plasmid multimerisation is not thought to be a factor in determining competition outcome in *umuDC*⁻ cells.

There may be a general reason for the loss of intensity of competition for all TR plasmids in *umuDC*⁻ cells, such as altered plasmid or cellular metabolism. Constitutive SOS induction or cellular filamentation have not been reported in *umuDC*⁻ strains.

During the 12-day incubations in the course of these experiments, the SOS response is likely to be induced several times as cultures repeatedly enter stationary phase. RecA-stimulated auto digestion of UmuD into UmuD' would occur. Thus translesion synthesis or inhibition of RecA-mediated DNA pairing (Rehraurer *et al.*, 1998) may play a role in TR replication or repair. It is possible that the absence of translesion synthesis in $(CCG/CGG)_{24}$ tracts is responsible for the loss of CCG/CGG_{24} plasmid competition in *umuD*⁻ compared to wild type cells. Future work may be able to address this specific question by sequencing TR plasmid DNA from wild type cells which have undergone SOS induction. If the frequency of point mutations in short TR tracts is greater than in vector pUC18 DNA, this may indicate that short TR tracts frequently provide substrates for error-prone translesion synthesis.

Summary

Following on from the negative results of the SOS-reporter assays carried out in chapter 5, the competition experiments carried out in this chapter were designed to establish whether TR tracts have any effects on the propagation of their host DNA molecule. The twelve day incubations in this study exposed the cells to both logarithmic growth and stationary phase growth several times, so that any effect in any phase of growth would be detected. Consequently, whilst clear differences in plasmid maintenance were observed, it is not directly apparent from this study which phases of culture growth allow mechanisms of plasmid loss or gain. As such, this study can only serve as a starting point for more specific experiments in which identical cultures could be maintained in stationary or exponential growth conditions.

The work performed here shows that length of the TR tract is by far the most significant factor in the ability of the plasmids tested here to compete with unmodified pUC18 control plasmids starting in the same cell. The ability of shorter repeat tracts ($CAG/CTG_{25,28}$ and CCG/CGG_{24}) to out-compete pUC18, whilst their longer brothers (CAG/CTG_{43}) lose to pUC18, is intriguing. There appears to be a critical length of TR tract, beyond which the TR tract becomes a liability for the maintenance of a *chi*-less plasmid in *E. coli*. It could be that longer TR tracts have the ability to form larger or more stable secondary structures, which inhibit replication to a greater degree, or are more prone to DSBs (perhaps as a result of hairpin cleavage).

One possible mechanism for the loss of mouse CAG/CTG_{43} is plasmid degradation by RecBCD activity initiated at DSBs. Jakupciak and Wells (1999) showed that genetic instabilities in $(CAG/CTG)_n$ repeats occurring by recombination needed repeat tracts of a minimum of 30 triplets, as well as functional *recA* and *recBC*. It is interesting to note that $(CAG)_{30}$ is also the threshold tract length for severe symptoms in myotonic dystrophy patients (Paulson and Fischbeck, 1996). In addition, the phenomenon of anticipation correlates with data demonstrating that short TR tracts in humans exhibit relatively small changes in length compared to the large ones exhibited by TR tracts greater than a threshold length of about 35 triplets.

Longer TR tracts might also inhibit replication either directly through the formation of secondary structures, or as a result of inhibited transcription elongation leaving stable DNA/RNA/transcription machinery complexes which act as replication fork road

blocks. Apart from directly reducing the rate of replication of mouse CAG/CTG_{43}, these structures may also induce an UmuD-dependent timed pause between stationary phase and exponential phase growth, which would favour segregants containing only pUC18 (Bowater *et al.*, 1996). However, there clearly must be other mechanisms of CAG/CTG_{43} loss, as pUC18 is still maintained preferentially in *umuDC*$^-$ cells. The reduced rate of short TR plasmid gain in this background is more difficult to explain, and casts doubts as to whether plasmid maintenance can readily be compared in wild type and *umuDC*$^-$ strains. One possibility is that translesion synthesis occurs at substrates present in short TR tracts, and results in a short primer suitable for standard DNA replication.

The relative gain of the short TR tract-bearing plasmids in wild type-cells was an unexpected but reproducible result. Their advantage in plasmid maintenance is most likely a result of an elevated frequency of replication priming, either from stable R-loop or D-loops formed in TR tracts, or as a result of translesion synthesis. This may suggest a replicative function enabling tolerance of short TR tracts in humans. It is difficult to imagine how D-loops generated during recombinational repair of DSBs or ss gaps would occur more often in short TR tracts than in longer TR tracts. R-loops on the other hand may exhibit length-dependent stability. Short R-loops may sometimes be able to act as primers for replication, but their small size means they are not stable enough to block unwinding by a replicative helicase in front of a replication fork, so do not inhibit replication. As such, short TR tracts may confer a replicative advantage to their host plasmids. Longer TR tracts on the other hand may be able to form longer and therefore more stable R-loops which do stall replication forks. The structures resulting from stalled replication do not allow sufficient *recA** formation or persistence to de-repress the SOS response, but may result in an UmuD-dependent timed pause resulting in a slower transition from stationary to exponential growth phase. This would confer a competitive advantage to daughter cells containing only pUC18 or deleted TR tracts. Unfortunately, the degree of plasmid segregation at the end of the competition studies was not established, and the intensity of competition with pUC18 control is reduced for both short and long TR plasmids in *umuDC*$^-$ cells. Non-the-less, the issue of stable R-loop formation is a crucial variable which must be addressed in future work, preferably by locating TR tracts outside transcribed areas of the plasmid.

Chapter 7: End-labelling experiments to determine trinucleotide repeat tract instability in different DNA replication, recombination, and repair mutants in *E. coli*.

Aims

The experiments presented so far in this thesis have focused specifically on recombination and replication in trinucleotide repeat tracts. The experiments within this chapter aim to determine whether deficiencies in specific recombination activities can be correlated to dynamic mutations within TR tracts. This was done by direct quantification of the instability of trinucleotide repeat tract length after introduction into recombination deficient strains. Instability was related to the purity of the dominant TR tract length in the starting material in order to determine the relative degree of instability in each mutant background. For comparison, strains bearing mutations in DNA replication, mismatch repair, and excision repair were also tested. In previous isolated reports, each of these repair pathways has been cited by others to affect TR instability. Whilst chapter 5 demonstrated that TR tracts used in this thesis do not induce the SOS response in *E. coli*, it is unknown whether any of the wide array of DNA-interacting activities regulated by this system have any effect on TR stability. Therefore strains constitutively induced or repressed for SOS response were also tested to rapidly screen the effects of up-regulating a wide variety of DNA-repair activities. In addition, the specific question of whether the reported ability of RecQ to unwind quadruplex structures in $(CCG/CGG)_n$ tracts actually reduces $(CCG/CGG)_n$ instability *in vivo*, was addressed by comparing wild type, *recQ*$^-$, *recJ* and *recQ*$^-$ *recJ*$^-$ strains.

Methods

In this chapter TR tracts of known length were examined for stability (the proportion of recovered TR tract still maintaining the original dominant tract size) in various backgrounds. The starting material for these studies were aliquots of the same *recA*$^-$ plasmid maxi-preps used in the other plasmid-based TR studies presented in this thesis. Two TR tracts were examined in both orientations: synthetic $(CCG/CGG)_{24}$ and mouse $(CAG/CTG)_{43}$.

100ng (1 µl) of plasmid DNA was used to transform 100 µl of $CaCl_2$-competent cells. Recovering cultures were plated on LB plates supplemented with ampicillin (100µg ml^{-1}). After a single overnight incubation, all of the colonies were rinsed from the plates and collected in 5ml LB. 1ml of this was centrifuged in a bench-top centrifuge for 1 minute to pellet the cells. 250µl of Qiagen buffer P1 was used to re-suspend the cells enabling TR plasmid purification by Minipreps.

20 µl of purified TR plasmids was digested for 2 hours at 37°C using 2 Units EcoRI to remove the TR tract. This DNA was end-labelled with [α-^{35}S] dATP, and 10µl samples run on 6% polyacrylamide gels for 2 hours to separate bands differing in length by a single triplet. The gels were dried for 45 min at 80°C. After overnight incubation in a Molecular Dynamics cassette, the radioactive images of these gels were scanned using a Storm860 scanner. Bands were quantified using Imagequant software.

From initial gels it was noted that the recovery of TR tracts varied between strains (perhaps as a result of plasmid copy number variation between strains, or different propensities for deletion of the TR tracts). Therefore, after an initial polyacrylamide gel to establish the recovery of end-labelled TR tract, a second polyacrylamide gel was run with the same samples adjusted for volume to give approximately an equal intensity of their dominant starting band. This improved comparison of gel bands by eye.

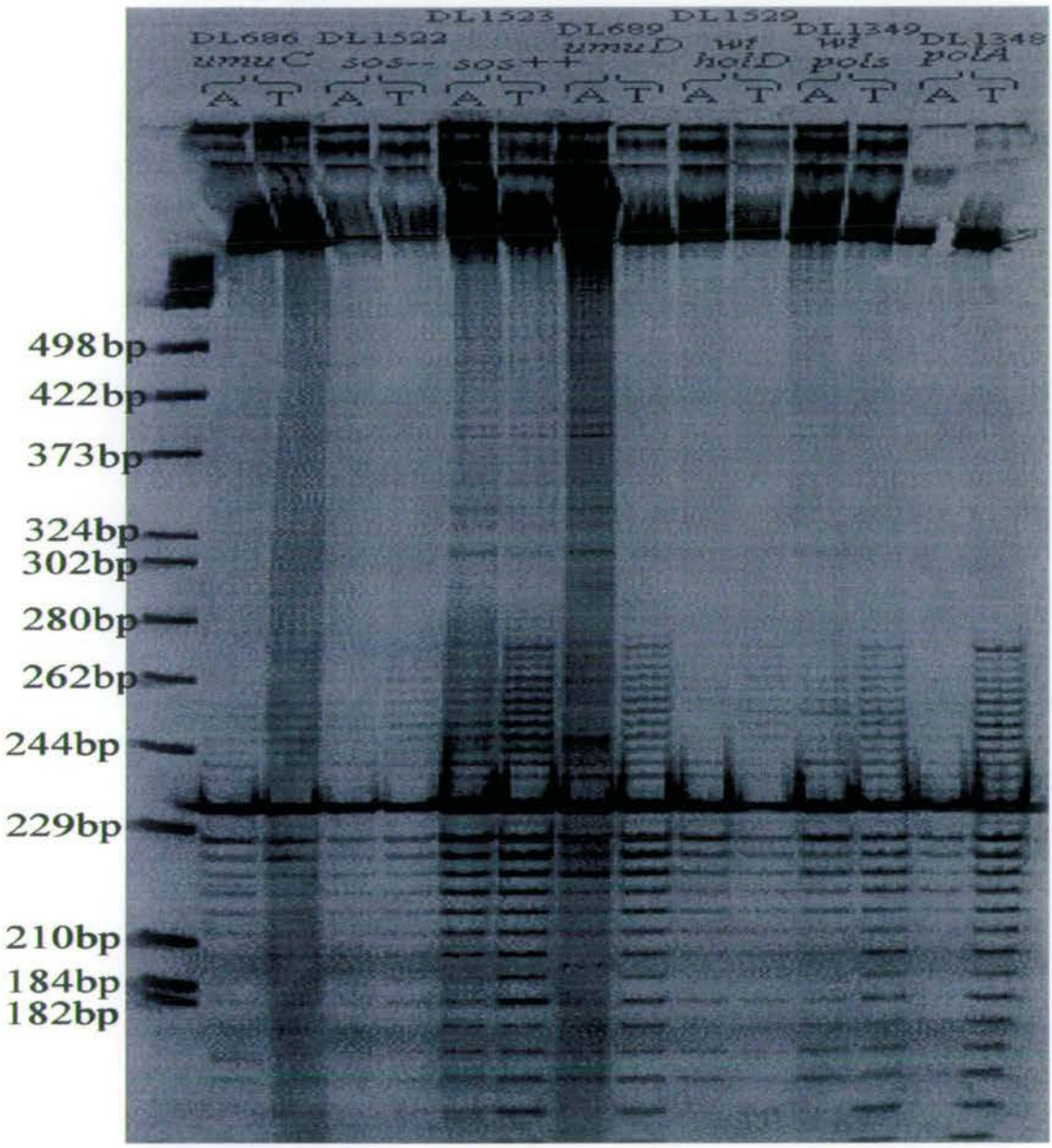
In *E. coli*, the process of transformation itself has been reported to produce some degree of instability within direct repeats and TR tracts (Hashem *et al.*, 2002). The deletion frequency of a 106bp inverted repeat in pBR325 was elevated by a factor of 2×10^5 following transformation into HB101, when compared with cells already holding the plasmid which were made competent. In an identical experiment, a $(CAG)_{76}$ tract was completely deleted 75-fold more frequently as a result of transformation. However, the same study found this transformation-mediated instability to be negligible for $(CAG)_{43}$, which is the longest TR tract length studied here.

It should also be noted that during the experiments presented in this chapter, cells would not have passed into stationary phase at any time. This would eliminate any selection for deletion products, which have previously been reported to result from passage through stationary phase growth (Bowater *et al.*, 1996).

Results

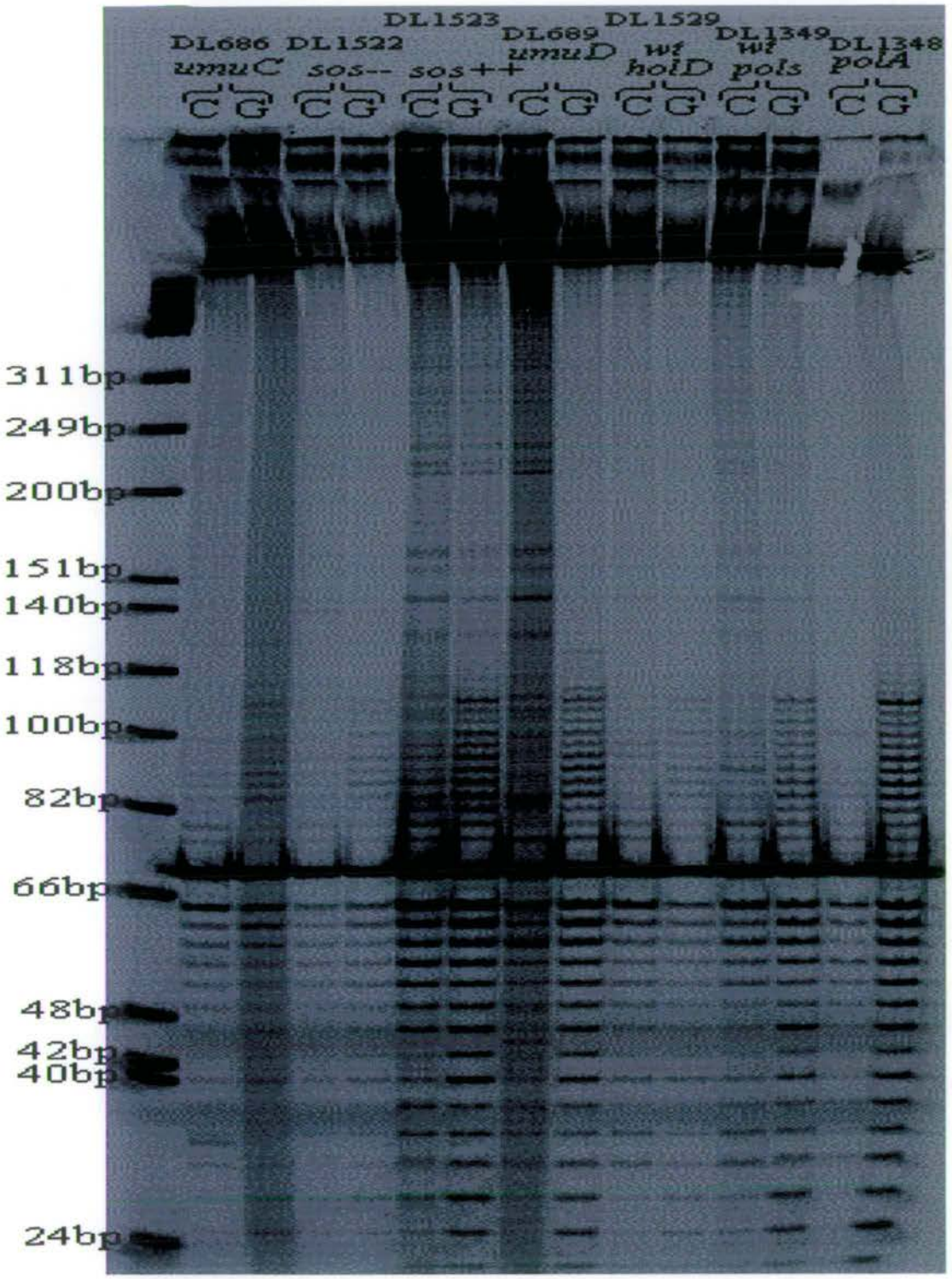
Two example gels are shown to illustrate gel clarity and the purity of the dominant starting bands of the four TR tract plasmids examined.

Figure 7.1: A polyacrylamide gel showing the effects of SOS induction and polymerase mutation on $(CAG/CTG)_{43}$ TR tract instability.



Various strains of *E. coli* carrying mutations in genes functioning in DNA-repair were transformed with plasmids carrying TR tracts (A = CAG_{43}, T = CTG_{43}), and individual colonies selected for overnight incubation in LB+Ampicillin. Plasmid mini-preps were performed, followed by S35 end-labelling of the excised EcoRI fragment carrying the TR tracts. The resulting samples were run on 6% polyacrylamide gels to enable resolution of TR tracts differing in length by single base triplets.

Figure 7.2: A polyacrylamide gel showing the effects of SOS induction and polymerase mutation on $(CCG/CGG)_{24}$ TR tract instability.



Various strains of *E. coli* carrying mutations in genes functioning in DNA-repair were transformed with plasmids carrying TR tracts (C = CCG_{24}, T = CGG_{24}), and individual colonies selected for overnight incubation in LB+Ampicillin. Plasmid mini-preps were performed, followed by S35 end-labelling of the excised EcoRI fragment carrying the TR tracts. The resulting samples were run on 6% polyacrylamide gels to enable resolution of TR tracts differing in length by single base triplets.

In almost every strain tested, the dominant starting band remains very well represented, so only a small proportion of TR tract copies experience changes in tract length. During quantification it was noted that triplet band intensity diminished with distance from the dominant starting band. Mechanistically this would suggest that changes in repeat tract length are either small, sequential, and cumulative, or that the probability of any given change in repeat length is approximately inversely proportional to that change in length. Significantly heavy bands were not observed away from the dominant starter band out-with this pattern, so no evidence was found to support the formation of secondary structures of specific length, or at least, these structures are not processed in an identical way every time they occur. In response to these trends, when interpreting the data obtained from these end-labelling experiments, analysis was restricted as to what degree particular mutants made TR tracts unstable, and whether they conferred a relative propensity for expansions or deletions.

Closer inspection of quantification data revealed that the intensity of expansion and deletion products was rather "flat" within tract lengths of +/- 13 triplets from the dominant starting band. This argues in favour of TR tract length changes occurring in isolated events in which the probability of any given change in repeat length is approximately inversely proportional to that change in length. This is in contrast to an alternative mechanism in which changes in repeat tract length are small, sequential, and cumulative. This latter model would predict approximately Poisson distribution curves for band intensity, as longer expansions would have to pass through several intermediate sizes of expansion as tract length gradually increased. In such a model, a small proportion of tract copies of each length (based on the high intensity of the dominant starting band after plasmid extraction) would subsequently experience instability, so there would be a rapid fall-off in band intensity.

Figure 7.3: The percentage of expansions, deletions and no net change to CCG24 tracts recovered from various backgrounds

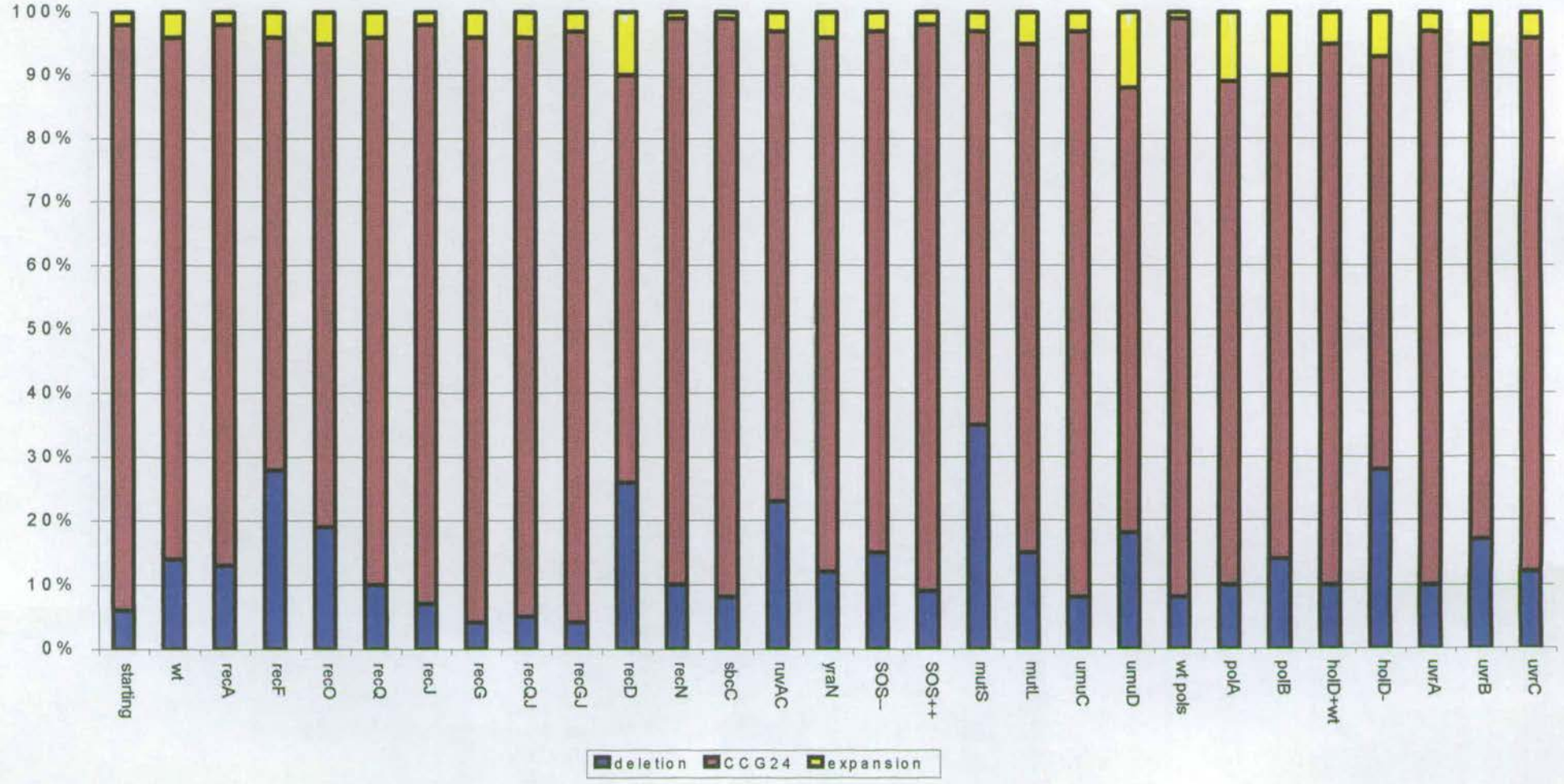


Figure 7.4: The percentage of expansions, deletions and no net change of CGG24 tracts recovered from various backgrounds.

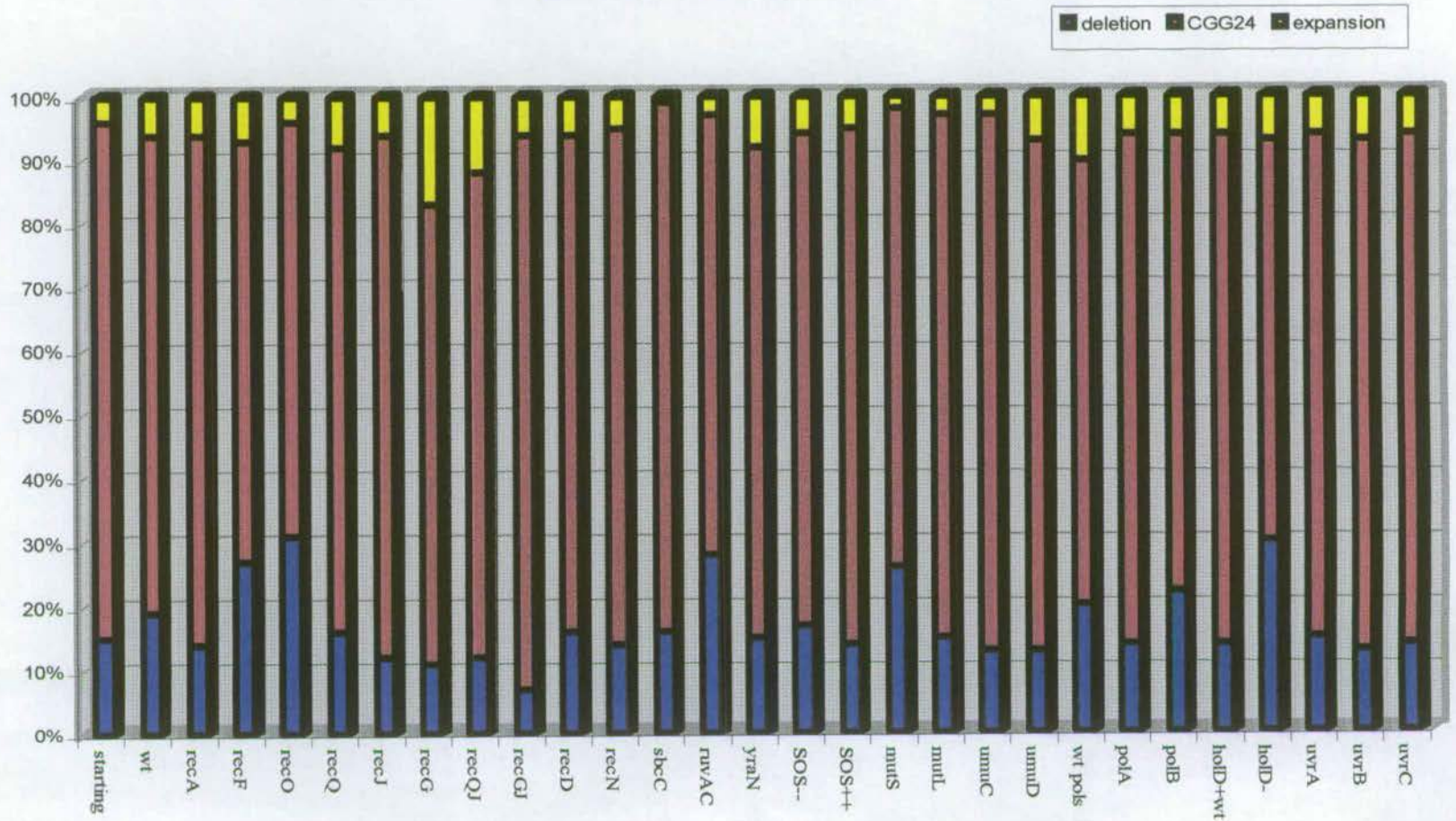


Figure 7.5: The percentage of expansions, deletions and no net change of CAG43 tracts recovered from various backgrounds

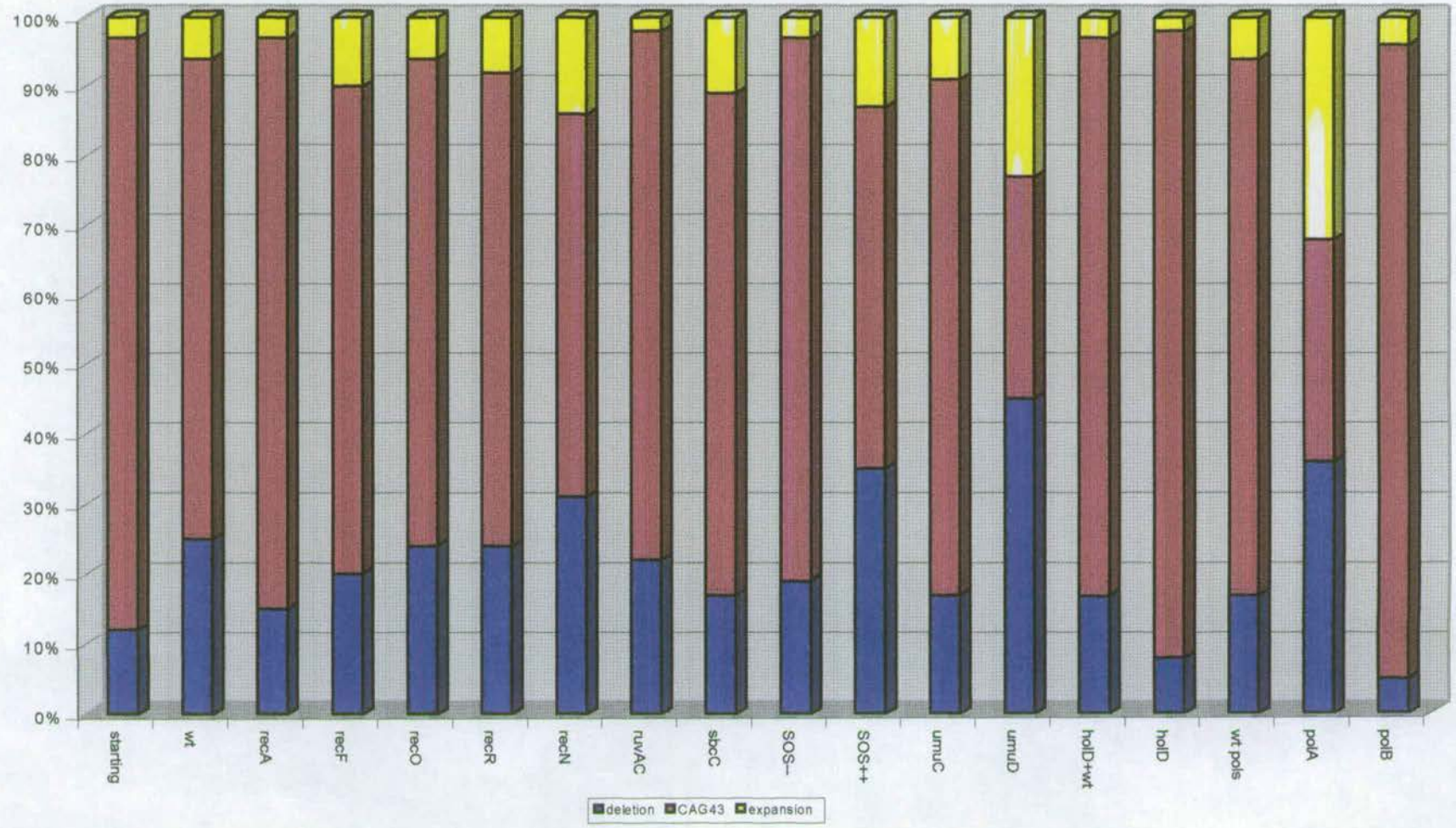


Figure 7.6: The percentage of expansions, deletions and no net change of CTG43 tracts recovered from various backgrounds

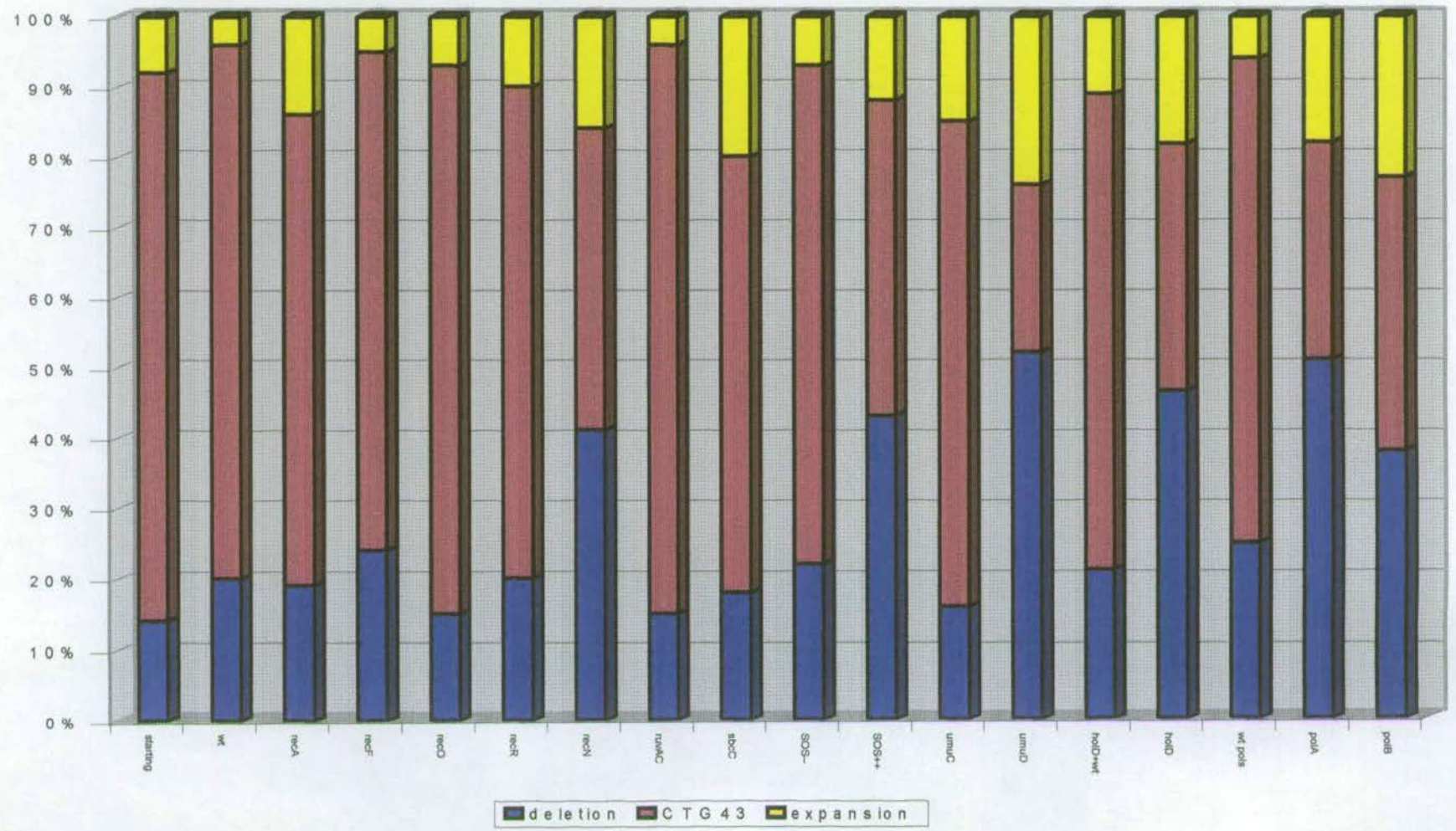


Table 7.1: Proportions of deleted, expanded, and constant TR tract lengths in DNA repair mutants.*

$(CCG)_{24}$

strain	genotype	deln	$(CCG)_{24}$	expn	Del/Exp
Start	*recA*⁻	6%	92%	2%	3.0
DL324	*wt*	14%	82%	4%	3.5
DL150	*recA*	13%	85%	2%	6.5
DL1368	*recF*	28%	68%	4%	7.0
DL1108	*recO*	19%	76%	5%	3.8
DL1301	*recQ*	10%	86%	4%	2.5
DL1096	*recJ*	7%	91%	2%	3.5
DL1077	*recG*	4%	92%	4%	1.0
DL1441	*recQJ*	5%	91%	4%	1.3
DL1440	*recGJ*	4%	93%	3%	1.3
DL1556	*recD*	26%	64%	10%	2.6
DL1106	*recN*	10%	89%	1%	10.0
DL515	*sbcC*	8%	91%	1%	8.0
DL1102	*ruvAC*	23%	74%	3%	7.7
DL1333	*yraN*	12%	84%	4%	3.0
DL1522	*SOS--*	15%	82%	3%	5.0
DL1523	*SOS++*	9%	89%	2%	4.5
DL1397	*mutS*	35%	62%	3%	11.7
DL386	*mutL*	15%	80%	5%	3.0
DL686	*umuC*	8%	89%	3%	2.7
DL689	*umuD*	18%	70%	12%	1.5
DL1349	*wt pols*	8%	91%	1%	8.0
DL1348	*polA*	10%	79%	11%	0.9
DL1350	*polB*	14%	76%	10%	1.4
DL1529	*holD+wt*	10%	85%	5%	2.0
DL1530	*holD-*	28%	65%	7%	4.0
DL699	*uvrA*	10%	87%	3%	3.3
DL698	*uvrB*	17%	78%	5%	3.4
DL695	*uvrC*	12%	84%	4%	3.0

$(CGG)_{24}$

strain	genotype	deln	$(CGG)_{24}$	expn	Del/Exp
Start	*recA*⁻	15%	81%	4%	3.8
DL324	*wt*	19%	75%	6%	3.2
DL150	*recA*	14%	80%	6%	2.3
DL1368	*recF*	27%	66%	7%	3.9
DL1108	*recO*	31%	65%	4%	7.8
DL1301	*recQ*	16%	76%	8%	2.0
DL1096	*recJ*	12%	82%	6%	2.0
DL1077	*recG*	11%	72%	17%	0.6
DL1441	*recQJ*	12%	76%	12%	1.0
DL1440	*recGJ*	7%	87%	6%	1.2
DL1556	*recD*	16%	78%	6%	2.7
DL1106	*recN*	14%	81%	5%	2.8
DL515	*sbcC*	16%	83%	1%	16.0
DL1102	*ruvAC*	28%	69%	3%	9.3
DL1333	*yraN*	15%	77%	8%	1.9
DL1522	*SOS--*	17%	78%	5%	3.4
DL1523	*SOS++*	14%	81%	5%	2.8
DL1397	*mutS*	26%	72%	2%	13.0
DL386	*mutL*	15%	82%	3%	5.0
DL686	*umuC*	13%	84%	3%	4.3
DL689	*umuD*	13%	80%	7%	1.9
DL1349	*wt pols*	20%	69%	10%	2.0
DL1348	*polA*	14%	80%	6%	2.3
DL1350	*polB*	22%	71%	6%	3.7
DL1529	*holD+wt*	14%	80%	6%	2.3
DL1530	*holD-*	30%	63%	7%	4.3
DL699	*uvrA*	15%	79%	6%	2.5
DL698	*uvrB*	13%	80%	7%	1.9
DL695	*uvrC*	14%	80%	6%	2.3

Mouse $(CAG)_{43}$

strain	genotype	deln	$(CAG)_{43}$	expn	Del/Exp
Start	*recA*	12%	85%	3%	4.0
DL324	*wt*	25%	69%	6%	4.2
DL150	*recA*	15%	82%	3%	5.0
DL1368	*recF*	20%	70%	10%	2.0
DL1108	*recO*	24%	70%	6%	4.0
DL1110	*recR*	24%	68%	8%	3.0
DL1106	*recN*	31%	55%	14%	2.2
DL1102	*ruvAC*	22%	76%	2%	11.0
DL515	*sbcC*	17%	72%	11%	1.5
DL1522	*SOS--*	19%	78%	3%	6.3
DL1523	*SOS++*	35%	52%	13%	2.7
DL686	*umuC*	17%	74%	9%	1.9
DL689	*umuD*	45%	32%	23%	2.0
DL1349	*wt pols*	17%	77%	6%	2.8
DL1348	*polA*	36%	32%	32%	1.1
DL1350	*polB*	5%	91%	4%	1.3
DL1529	*holD+wt*	17%	81%	3%	5.7
DL1530	*holD*	8%	90%	2%	4.0

Mouse $(CTG)_{43}$

strain	genotype	deln	$(CTG)_{43}$	Expn	Del/Exp
Start	*recA*	14%	78%	8%	1.8
DL324	*wt*	20%	76%	4%	5.0
DL150	*recA*	19%	67%	14%	1.4
DL1368	*recF*	24%	71%	5%	4.8
DL1108	*recO*	15%	78%	7%	2.1
DL1110	*recR*	20%	70%	10%	2.0
DL1106	*recN*	41%	43%	16%	2.6
DL1102	*ruvAC*	15%	81%	4%	3.8
DL515	*sbcC*	18%	62%	20%	0.9
DL1522	*SOS--*	22%	71%	7%	3.1
DL1523	*SOS++*	43%	45%	12%	3.6
DL686	*umuC*	16%	69%	15%	1.1
DL689	*umuD*	52%	24%	24%	2.2
DL1349	*wt pols*	25%	69%	6%	4.2
DL1348	*polA*	51%	31%	18%	2.8
DL1350	*polB*	38%	39%	23%	1.7
DL1529	*holD+wt*	21%	67%	11%	1.9
DL1530	*holD*	46%	35%	18%	2.6

* deln is deletion. expn is expansion. Del/Exp is the ratio of deletions to expansions.

Table 7.2: A comparison of average instability values for each TR plasmid.

	deletion	CCG_{24}	expansion	deletion:expansion
starting material	6%	92%	2%	3.0
average of wt	11%	86%	3%	3.2
average from all strains tested in common	15%	80%	5%	3.0
	Deletion	CGG_{24}	Expansion	deletion:expansion
starting material	15%	81%	4%	3.8
average of wt	18%	75%	7%	2.4
average from all strains tested in common	22%	73%	5%	4.5
	Deletion	CAG_{43}	Expansion	deletion:expansion
starting material	12%	85%	3%	4.0
average of wt	20%	76%	5%	3.9
average from all strains tested in common	23%	67%	10%	2.2
	Deletion	CTG_{43}	Expansion	deletion:expansion
starting material	14%	78%	8%	1.8
average of wt	22%	71%	7%	3.1
average from all strains tested in common	30%	55%	15%	2.0

average of wt = average of wt AB1157, the wt strain in which polymerase mutations were isolated, and the wt strain in which the *holD* mutation was isolated.

average from all strains tested in common = average of strains tested with all four TR plasmids (w.t., *recA , recF, recO, recN, sbcC, SOS-, SOS+, umuC, umuD, ruvAC, holD, polA, polB*).

General observations

In all cases, passage of the TR plasmid through any strain increased the instability of that repeat. Even in the *recA*$^-$ strain (the same recombination deficient strain that the original starting plasmid maxi-preps were isolated from), all of the plasmids display elevated instability, except $(CAG)_{43}$, which remain unchanged. In all of the strains tested, the $(CAG/CTG)_{43}$ repeats appear to be more unstable than the $(CCG/CGG)_{24}$ repeats. The greatest levels of instability were observed in $(CTG)_{43}$ tracts in the *polA*$^-$ and *holD*$^-$ strains. None of the TR sequences in either orientation displayed a greater propensity for expansion than for deletion in any strains tested.

Artificial CCG_{24}

CCG_{24} was very stable in wild type cells. Elevated frequencies of CCG_{24} deletion were seen in the following recombination mutants: *recF*⁻, *recD*⁻, and *ruvAC*⁻. Deleted CCG_{24} tracts were also observed with greater frequency in *mutS*⁻ cells and *holD*⁻ cells. CCG_{24} stability resembled wild type values in *sbcC*⁻ cells, and cells either constitutively activated or constitutively repressed for the SOS response. Whilst *uvrB*⁻ cells may have displayed slightly higher than wild type rates of deletion, *uvrA*⁻ and *uvrC*⁻ strains displayed no change from wild type. Large changes in the rate of CCG_{24} amplification were not observed in any of the strains tested, though slightly elevated rates of amplification were observed in *recD*⁻, *umuD*⁻, *polA*⁻, and *polB*⁻ strains.

Artificial CGG_{24}

In almost every strain tested, CGG_{24} repeats proved to be more unstable than their CCG_{24} counterparts (the possible exceptions being in the *mutS*⁻ and *mutL*⁻ strains). Compared to wild type, elevated frequencies of CGG_{24} deletion were seen in the following recombination mutants: *recF*⁻, *recO*⁻ and *ruvAC*⁻. Elevated deletion rates were also observed in *mutS*⁻, and *holD*⁻ strains. A large increase in TR expansion was seen in *recG*⁻ cells, 17% being the largest increase in the proportion of $(CCG)_{24}$ or $(CGG)_{24}$ tracts observed in any strains.

Mouse CAG_{43}

Rates of $(CAG/CTG)_{43}$ deletion were enhanced in *recN*⁻, *umuD*⁻, and *polA*⁻ strains and in cells constitutively induced for the SOS response. The elevated $(CAG/CTG)_{43}$ instability observed in *umuD*⁻ and *polA*⁻ strains was also due to an elevated rate of expansion compared to wild type. Mutations in *recF, recO* and *recR* had little effect on wild type levels of instability.

Mouse CTG_{43}

This TR tract was found to be the most unstable of all those tested. Between 16% and 52% of all of the TR tracts recovered from the strains tested here exhibited deletions.

In particular, mutations in *recN*$^-$, *umuD*$^-$, *holD*$^-$, *polA*$^-$, and *polB*$^-$ greatly increased the propensity for deletion of $(CTG)_{43}$, as did constitutive induction of the SOS response. The rate of CTG_{43} expansion in wild type cells was only elevated in *umuD*$^-$, and *polB*$^-$ strains, and even these increases were modest. Mutations in *recF, recO* and *recR* had little effect on wild type levels of instability, as did constitutive repression of the SOS response.

Discussion

General stability observations

CCG_{24} tracts proved to be extremely stable. CGG_{24} tracts were slightly less so. CAG_{43} and CTG_{43} tracts were less stable. It is perhaps significant that the longer TR tracts tested were above the disease threshold length of approximately 35 repeats in expressed polyglutamine diseases. The elevated instability found here above this length may be correlated to anticipation in human pedigrees, in which generations following an expansion above a threshold length are expected to develop more severe symptoms of the disease as tract length increases in successive generations. In support of this, sperm typing of CAG tract instability in the androgen receptor (responsible for SBMA) in approximately 4,300 sperm showed that individuals with 28-31 repeats had 4.4 times greater rates of dynamic mutation than individuals carrying alleles with 20-22 repeats (Zhang *et al.*, 1994).

It would seem that in all of the TR plasmids examined here, there appears to be a general bias towards deletion rather than expansion of the TR tracts. In germline mutation events monitored by sperm typing, contractions also outnumbered expansions, but by a ratio of nine to one (Zhang *et al.*, 1994). Deletions in $(CAG)_{39}$ were observed to occur more often than expansions in yeast (Richard and Paques, 2000; Balakumaran *et al.*, 2000). Previous work in *E. coli* has also found that deletions occur more frequently than expansions (Kang *et al.*, 1995). It is not clear whether this effect results from a preferred mechanism of instability leading to deletions, or whether TR instability results in some deletion products, which are subsequently selected for in the population of cells. If selection for deletion products does occur, it must operate in exponential phase growth since those are the only conditions to which the cells were exposed in this experiment.

Orientation effects on TR tract instability

In wild type cells, $(CGG)_{24}$ present on the lagging strand was approximately 10% less stable than $(CCG)_{24}$ on the lagging strand of replication. Also in wild type cells, $(CTG)_{43}$ was approximately 20% less stable than $(CAG)_{43}$ on the lagging strand of replication. Single stranded $(CTG)_n$ tracts form more stable hairpins than $(CAG)_n$ tracts of the same length (Gacy *et al.,* 1995), and $(CGG)_n$ repeats form more stable hairpin and quadruplex structures than $(CCG)_n$ tracts of the same length (Mitas, 1997; Gacy and McMurry, 1998). All of these secondary structures would form in ssDNA, which is assumed to be more common on the lagging strand due to discontinuous synthesis. Although the TR tracts examined in this study are not of the same triplet sequence, and so not directly comparable, it is also known that longer TR tracts form more stable hairpins (Gacy and McMurry, 1998). Similarly, palindromes known to form into secondary structures have previously been shown to exhibit a preference for deletion when present on the lagging strand of DNA replication (Pinder *et al.,* 1998). Therefore, all of the results presented here are consistent with secondary structure formation occurring predominantly on the lagging strand of DNA replication, and directly mediating TR tract instability, predominantly by elevating frequencies of triplet deletion to give contracted TR tract products.

Further support for this hypothesis comes from observations made by others which suggest that the orientation in which TR tracts are present with respect to a defined replication origin significantly affects tract stability. In yeast, $(CTG)_{130}$ tracts are ten times more unstable when present on the lagging strand of replication (Maurer, O'Callaghan and Livingston, 1996; Freudenreich, Stavenhagen and Zakian, 1997). Also in yeast, deletions and expansions were more frequent when $(CGG)_{81}$ and $(CGG)_{190}$ were present on the lagging strand of replication, with contractions occurring up to thirty times more frequently than expansions (Balakumaran, Freudenreich, and Zakian, 2000). This last study was particularly interesting because the research identified a 100-200 fold elevated recombination frequency at the TR tracts independent on orientation (compared to no-tract control), dependent of gene products known to be involved in yeast DSBR. It would therefore seem that "fragile" $(CCG/CGG)_n$ tracts cause DSBs, which must be repaired by recombination. Thus

$(CCG/CGG)_n$ tract instability was shown to be orientation dependent, but $(CCG/CGG)_n$ tract fragility was shown to be orientation independent. By comparing earlier studies performed in the same laboratory, recombination measured by generation of resistance to 5-fluoro-orotic acid at $(CGG)_{160}$ tracts occurred at 3.4×10^{-3} events per cell generation, where as $(CTG)_{250}$ tracts induced recombination at 2.4×10^{-3} events per cell generation. Therefore, in yeast, a shorter (CGG) repeat tract is required to give the same rate of breakage as a longer (CTG) repeat tracts.

The dependence of TR tract instability on the orientation of replication has been shown in a variety of model organisms, and is compelling evidence for the involvement of replication in the mechanism of TR instability. Also consistent with hairpin formation in the mechanism of TR tract instability is the observation that the absence of single-stranded-DNA-binding protein (which prevents secondary structure formation when bound to ssDNA) leads to an increase in the frequency of large deletions within the TR tracts (Rosche *et al.*, 1996).

Recombination mutants

The fact that TR tract instability still occurs in *recA*$^-$ cells at levels highly comparable to wild type discourages faith in models proposing recombination as the dominant form of TR tract instability. Mutations in *recF, recO,* and *recR* do not appear to alter $(CAG/CTG)_{43}$ instability from wild type levels, though there may be a slight increase in $(CCG/CGG)_{24}$ instability in *recF*$^-$, *recO*$^-$, and *ruvAC*$^-$ strains. Indeed, these CCG/CGG_{24} results would suggest that improper regulation of homologous recombination (such as progression of RecA filament into areas of dsDNA, or failure to resolve Holliday junctions in a co-ordinated fashion) may mediate some degree of instability which is not observed in the absence of RecA. Unfortunately a *recD*$^-$ strain was not tested with $(CAG/CTG)_{43}$ tracts in this study, so the crucial question of recombinational DSBR could not be tested on these *chi*-less plasmids bearing long TR tracts. However, the shorter TR tracts were tested in this strain. The $(CGG)_{24}$ tract did not display any differences to the instability observed in wild type cells. However, the $(CCG)_{24}$ tract displayed approximately five-fold increases in the recovery of both expanded and deleted tracts. It would seem that DSBR may affect the instability of TR tracts, but more experiments are required to verify this.

Although the original maxi-preps extracted the starting plasmid samples from a *recA*$^-$ strain, transformation back into the *recA*$^-$ strain decreased the proportion of dominant starter tract length. The recovery of less dominant starting tract length equates to TR tract instability either during the transformation process, or as a result of continued propagation through the *recA*$^-$ strain. However, transformation-mediated instability was previously found to be negligible for plasmids carrying $(CAG)_{43}$ tracts (Hashem *et al.*, 2002). It therefore seems that the instability observed in these experiments is a function of how many generations the TR tracts have passed through. One process which these TR tracts have most certainly encountered in proportion to the time spent in host cells, is DNA replication, and it was on the basis of these *recA*$^-$ results that this investigation was pursued in replication mutants.

Previous work on 787bp direct repeats has estimated that only 20% of rearrangements occurred in a RecA-independent fashion (Morag, Saveson and Lovett, 1999), since expansions and deletions were reduced five-fold in a *recA*$^-$ strain. However, these direct repeats would not be expected to be subject to replication slippage at rates as high as TRs because of their more limited capacity for homologous binding, so there would be expected to be greater levels of RecA-independent instability within TR tracts. Other work has suggested that instabilities at CAG/CTG repeats occur by RecA-mediated gene conversion (Jakupciak and Wells, 1999), which results from non-reciprocal homologous recombination. It is interesting that perfectly pure repeats are known to be more unstable than interrupted repeats and the introduction of point mutations in repeat tracts has been found to reduce the rate of expansion (Jakupciak and Wells, 2000). This is consistent with a RecA-mediated mechanism of instability, which would be inhibited by complementary strands slipping and producing mismatches (targets for MutS binding and subsequent inhibition of RecA filament-mediated strand exchange). Indeed, the loss of single nucleotide polymorphisms in stable, normal human TR alleles has been correlated with subsequent expansion of the TR tract (Gunter *et al.*, 1998; Kunst *et al.*, 1994; Eichler, 1994).

According to Jakupciak and Wells (1999 and 2000), recombination is one mechanism which can mediate TR tract instability. They used a two-plasmid system in which the replication dependent generation of a double strand break in the TR tract of one

plasmid is repaired using a larger repeat (present on another plasmid) as a template. They observed that recombination was the dominant form of expansion for long repeats, and the mechanism was one of gene conversion. No effect of TR orientation was observed on inter-plasmid recombination in this system, which they cited as further evidence for a recombination-mediated mechanism. The presynaptic event of homologous pairing might occur along any part of the exposed trinucleotide repeat tract, leading to unequal crossing over. In this model, the location of strand invasion would determine the size of expansion or deletion formed (Jakupciak and Wells, 1999). Plasmids with CAG/CTG repeat tracts as long as 175 triplets were used to prime recombination with plasmids carrying 36 direct repeats of CAG/CTG triplet, and the ratio of expansions to deletions was found to be as high as 100:1. However, they only observed expansions for TR tracts longer than Okazaki fragments (see possible recombination initiation at unprocessed Okazaki flaps, later). This contrasts with numerous studies in yeast and recombination deficient *E. coli*, which have determined expansions to be far less prevalent than deletions.

Kang *et al.*, (1996) used a CTA triplet interrupting $(CTG)_{175}$ as a polymorphism to demonstrate that large expansions occurred in single events, rather than a series of small steps. No expansions were observed in *recA*$^-$ cells by sequence analysis (consistent with a recombination mechanism) (Jakupciak and Wells, 1999). The experiments performed in this chapter did not detect large expansions, but did not use longer TR tracts which could act as templates for gene conversion events. The small TR tract size used in the experiments presented in this thesis may not be suited to examination of instability mediated by gene conversion, since the minimal strand length for efficient recombination in *E. coli* is reported to be 90b (Shen and Huang, 1986). RecA filament would therefore have to spread beyond the TR tract, and into flanking sequences, which could anchor the recombination event (preventing mis-pairing) to maintain the original tract size. In contrast to the RecA-dependent mechanism of TR tract expansion mentioned above, other studies have shown that *recA* is not required to generate deletions in TRs (Bowater *et al.*, 1996), and it would seem that it is this mechanism of TR instability that has been detected in the experiments presented here.

By comparing the effects of recombination deficient strains on TR plasmid multimerisation (chapter 4), and TR tract instability, it is clear that specific mutants which reveal TR tract-stimulated recombination, do not appear to induce TR tract instability during this process. In support of this, Jakupciak and Wells (1999) could not detect expansions in a $(CTG)_{36}$ tract in wild type AB1157, despite the formation of homodimers and homotrimers in their pUC19-derived vector. This would also suggest that the method of multimer formation in wild-type cells is independent of the method of TR expansion.

Secondary structure processing

Quadruplex unwinding by RecQ

Previous *in vitro* work has highlighted the ability of RecQ and its eukaryotic homologues to unwind DNA secondary structures formed in $(CCG/CGG)_n$ TR tracts, allowing progression of DNA synthesis through blocks which previously stalled replication (Kamath-Loeb *et al.*, 2001; Sun, Bennett and Maizels, 1999; Sun *et al.*, 1998). Any *in vivo* consequences of this reaction to $(CCG/CGG)_{24}$ instability were examined by specifically comparing $(CCG/CGG)_{24}$ extracts from populations of transformed *recQ*, *recJ*, and *recQJ* cells. Similar patterns of $(CCG/CGG)_{24}$ instability would suggest RecQ activity is only relevant to standard recombination functions in which DNA unwinding by RecQ is co-ordinated with the ss exonuclease activity of RecJ on the emerging 5' strand. Elevated instability in a *recQ* strain compared to a *recJ* strain would suggest that RecQ has a distinct role separate from RecJ, such as unwinding secondary structures which elevate $(CCG/CGG)_{24}$ instability.

The absence of *recQ*, either in otherwise wild type cells, or in combination with *recJ* did not cause significant elevations in $(CCG)_{24}$ or $(CGG)_{24}$ tract instability. Any variation between the single and double mutants tested was considered to be within experimental error. As such, these experiments indicate quadruplex unwinding by RecQ may not be a significant suppresser of instability in $(CCG/CGG)_{24}$ tracts. This may be because it is a redundant pathway, which can be substituted by, for example, hairpin processing. However, it may also be that quadruplex structures are not stable

in TR tracts of this length, so do not offer sufficient resistance to replication forks to cause fork stalling and the requirement for a dedicated unwinding activity. Finally, the $(CCG/CGG)_{24}$ tracts proved to be relatively stable in all of the strains tested, so either a more sensitive assay or longer $(CCG/CGG)_n$ tracts would perhaps be required to detect any change in tract stability in these backgrounds.

Hairpin cleavage by SbcCD

Hairpin cleavage by SbcCD is a crucial step in many models of TR tract expansion and deletion, not only because it directly removes a length of TR tract on one strand, but also because the product of this reaction is a recombination substrate. Previous studies have suggested that SbcC expression abolishes $(CTG)_{150\text{-}200}$ amplification in SURE cells (Sarkar *et al.*, 1998). The experiments performed in this chapter did detect a greater proportion of $(CAG)_{43}$ and $(CTG)_{43}$ tract expansions in *sbcC*$^-$ cells, with this effect being more pronounced in $(CTG)_{43}$ tracts. It is interesting to note the overall tract stability remains comparable to wild type levels for $(CAG)_{43}$, but there are decreased levels of deletion, which correspond to the gain in expansion products.

$(CCG)_{24}$ and $(CGG)_{24}$ tracts proved to be more stable in *sbcC*$^-$ cells than in wild type, remaining almost unchanged from the initial plasmid transformation samples. However, as mentioned above in the RecQ analysis, $(CCG/CGG)_{24}$ tract instability in wild type cells was not sufficiently high to obtain a reliable estimate of the contribution of SbcC to instability in these repeats, though they would suggest some effect. Again, a more sensitive assay must be sought.

SOS induction

Although the results from chapter 5 show that TRs do not induce the SOS response, some of the tracts tested here do behave differently in SOS-induced cells. Permanent induction of the SOS response had no effect on $(CCG)_{24}$ tracts, which gave results highly comparable to both wild type and constitutively un-induced SOS response. Permanent induction of the SOS response may have slightly elevated $(CGG)_{24}$ instability. $(CCG)_{24}$ and $(CGG)_{24}$ tract instability remained unchanged in cells unable to induce the SOS response. $(CAG)_{43}$ and $(CTG)_{43}$ tracts also displayed approximately wild type levels of instability in cells unable to induce the SOS response. However, a

comparatively strong destabilising effect on $(CAG)_{43}$ and $(CTG)_{43}$ tracts was observed when they were introduced into cells constitutively induced for the SOS response. This reduced stability occurs via an increase in the occurrence of both expansions and deletions. The results for $(CCG/CGG)_{24}$ and $(CAG/CTG)_{43}$ tract instability in cells constitutively induced for the SOS response are very similar in *recN*⁻ cells, which have also been reported to exhibit constitutive SOS induction (Dunman *et al.*, 2000).

A *lexA(Def)* mutation leading to constitutive SOS induction has previously been found to increase by 8-22 fold the instability of $(AC/TG)_{18\text{-}51}$ tracts inserted into the *E. coli* chromosome (Morel *et al.*, 1998). This increase in dinucleotide repeat tract instability was also observed in *polA*⁻ and *uvrD*⁻ mutants, which also induce the SOS response. The increase in dinucleotide repeat tract instability in SOS-induced cells was seen to be independent of RecA function. This work also found that *recA*⁻ and *sbcD*⁻ mutations had no effect on dinucleotide repeat instability in otherwise wild type cells. It was suggested that the mechanism of this SOS-induced dinucleotide instability may be a result of modified β sliding clamps elevating the frequency of DNA Pol III slippage. Interestingly, the spontaneous deletion of palindromes in plasmid pBR325 has also been observed to be increased by SOS induction (Balbinder *et al.*, 1993).

DNA damage checkpoint activation

$(CAG)_{43}$ and $(CTG)_{43}$ instability was highest in *umuD*⁻ cells out of all those tested. There was a strong preference for deletion in these tracts. $(CCG)_{24}$ and $(CGG)_{24}$ tracts on the other hand were far more stable, and exhibited instability values far more comparable to those observed in wild type cells. All of the TR tracts tested were shown to be far more unstable in *umuD*⁻ cells than in *umuC*⁻ cells. This confirmed that translesion synthesis by $UmuD_2$'C was not the mechanism of the instability observed in *umuD*⁻ cells. Bowater and Wells (1996) observed that cells containing long (CAG/CTG) tracts were subject to an extended lag phase of *E. coli* growth. Subsequent work demonstrated that this extended lag phase is the result of an UmuD-dependent DNA damage checkpoint, capable of initiating a timed pause in replication (Opperman *et al.*, 1999). The extended lag phase of the longer TR tracts gave a growth advantage to cells bearing deleted TR tracts, which led to a dramatic increase

in the frequency of deletions recovered. It is not thought that cells in the experiments presented here would have encountered a stationary to exponential growth phase transition, so wild type cells containing long TR tract plasmids would not be subject to extended lag phases. However, the elevated deletion and expansion rates observed in *umuD*$^-$ cells would suggest that the UmuD checkpoint is active under these experimental conditions, and functions to prevent long TR tract instability. Indeed, using cultures which had not passed through stationary phase, Bowater and co-workers (1996) observed a correlation between the proportion of recovered plasmids containing deletions, and the number of bacterial generations the plasmids were grown through. $(CAG/CTG)_{175}$ displayed an orientation effect for this. More deletions were observed in plasmids recovered from faster growing cultures than slower growing cultures (dictated by temperature, shaking, and medium richness). It is unclear whether this was a result of an apparent growth advantage of bacteria harbouring deleted TR tracts, or whether another mechanism, such as replication slippage occurred at a constant rate down the generations.

The greater instability observed in $(CTG)_{43}$ compared to $(CAG)_{43}$ tracts may point to a role of DNA secondary structure in the mechanism activating this replication pause. Additionally, the relative stability of $(CCG)_{24}$ and $(CGG)_{24}$ tracts in *umuD*$^-$ cells would suggest that these repeat tracts do not trigger the DNA damage checkpoint. Since these shorter TR tracts exhibit very little instability in all of the strains they were tested in, and showed little response to the removal of activities which resolve secondary structures, it would seem that they are less capable of forming secondary structures than their longer $(CAG/CTG)_{43}$ counterparts. Results from the competition studies in chapter 6 showed that the loss of CAG/CTG_{43} plasmids compared to pUC18 control was not as pronounced in *umuDC*$^-$ cells, suggesting that at least some of the problems they caused wild type cells were attributable to activation of an UmuD-dependent timed pause. Further work is required to establish what kind of conditions or substrates may activate this DNA damage checkpoint.

Mismatch repair

Small deletions and amplifications of a single triplet repeat were observed in the *mutS*⁻ mutant tested with CCG/CGG_{24} plasmids in this background. These +1/-1 changes in TR length have previously been observed by others in many systems (see MIN in introduction chapter1), and confirm the sensitivity of this assay. They are thought to result from an inability to repair small DNA loops of 3 bases, and confer reduced stability to all microsatellite sequences. Consequently *mutS*⁻ cells displayed an elevated instability of the dominant band of the $(CCG/CGG)_{24}$ plasmids tested in this study. In each case, there was found to be an approximately ten-fold bias towards deletion rather than expansion products. This would suggest that mismatches and small DNA loops of three base pairs occur frequently on the template strands of DNA synthesis, allowing the formation of deletions in the nascent strands through replication slippage. *mutS*⁻ mutations have also been observed to increase dinucleotide repeat tract instability by 26 to 160 fold (Morel *et al.*, 1998).

Polymerase mutations

Mutations in *polA* and *polB* have a slight decrease in CCG/CGG_{24} stability. This effect is more pronounced in CAG/CTG_{43} tracts, which are very unstable in *polA*$^-$ cells due to elevated rates of deletion.

polA$^-$

polA encodes DNA polymerase I, which is thought to function in Okazaki fragment processing, as well as UV and chemical mutagen repair by nick translation (Sharma and Smith, 1987). The 5' nuclease domain of DNA Polymerase I has structural and functional homology to the 5'→3' exonuclease activities of *S. cerevisiae* Rad27p and human FEN1 (Flap Endonuclease), which cleave 5' flaps generated by displacement synthesis as a DNA polymerase reaches the 5' end of an Okazaki fragment (Robins *et al.*, 1994, Harrington and Lieber, 1994, and Wu *et al.*, 1996). *polA*$^-$ strains are defective in Okazaki fragment processing and are dependent on RecBCD and RecA, as they generate many DSBs (Miguel and Tyrrell 1986, and Morel *et al.*, 1998), which may also explain the de-repression of the SOS response observed in *polA*$^-$ strains. DNA polymerase I also functions in the initial DNA replication of ColE1 type plasmids (including pUC18) by extending DNA replication 200 to 400 nucleotides downstream of the folded RNA primer. This reaction is believed to allow DNA Pol III access to the 3'-OH end, which would not otherwise be possible due to steric hindrances attributable to the folded DNA/RNA primer.

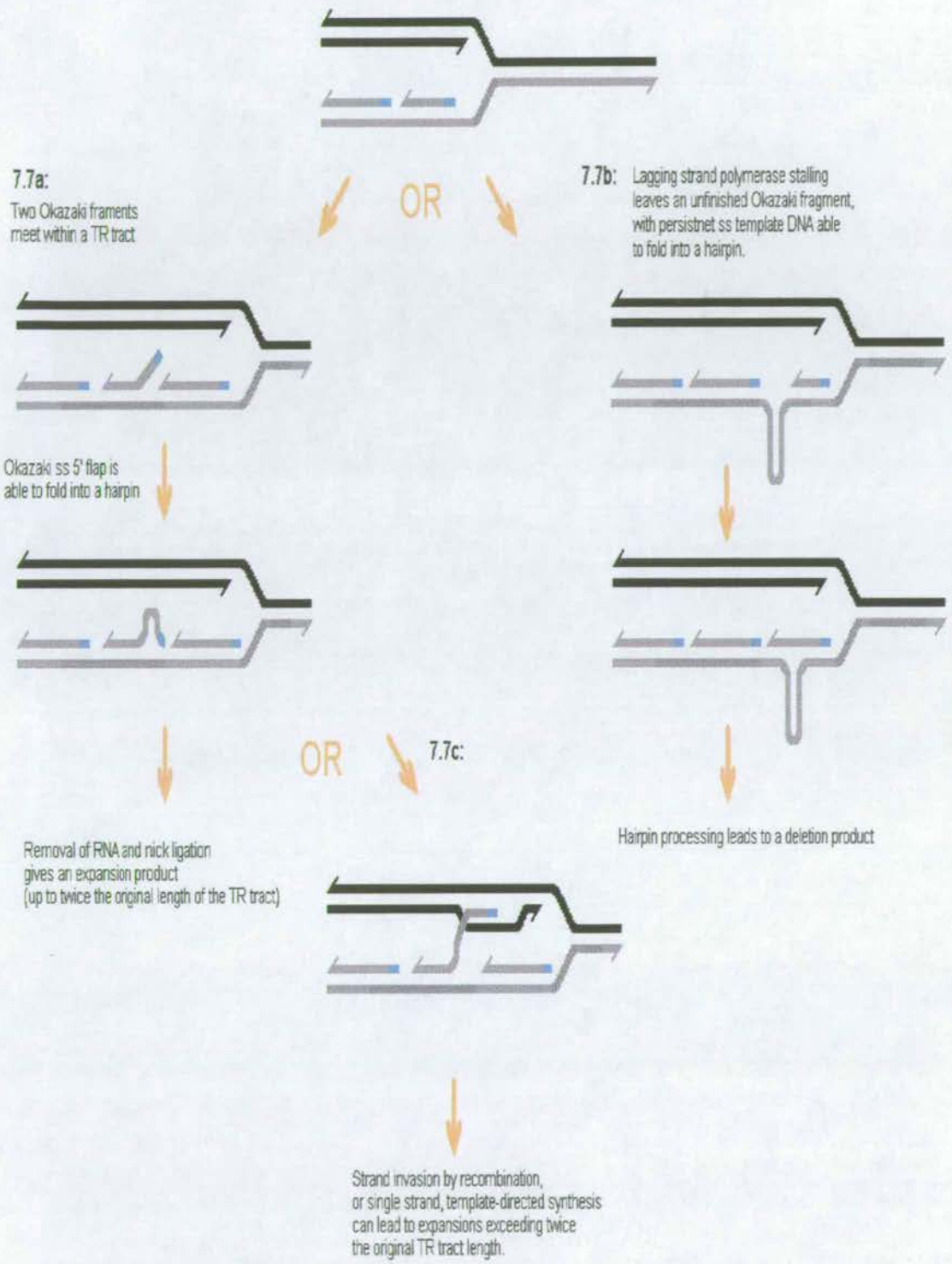
Figure 7.1 demonstrates several ways in which TR tract instability may result from defective flap processing. Expansions (Figure 7.1a) could result from the lagging strand polymerase displacing a ss polynucleotide flap from the 5' end of the previous Okazaki fragment, which are fixed in the absence of effective flap processing. *In vitro* work by Spiro and co-workers (1999) demonstrated that TR tracts capable of forming hairpins prevent FEN-1 access to the 5' end of the previous Okazaki fragments. Thus hairpin formation in the displaced 5' end of an Okazaki fragment can inhibit flap cleavage. However, this mechanism could only account for TR tract amplifications of up to twice the original TR length. Larger expansions may be possible if the 5' flap is able to invade a homologous region of duplex, establish a D-loop, and initiate

replication (Figure 7.1c). Clearly, this reaction would most likely be mediated by homologous recombination. Support for this comes from Δ*rad27* yeast, which produce a large proportion of expansion products exceeding twice the original length of the TR tracts (Spiro *et al.,* 1999). The recombination connection is reinforced by the fact that double mutants for *rad27* and any single enzyme involved in yeast homologous recombination (RAD1, RAD50, RAD51, RAD52, RAD54, RAD55, RAD57, MRE11, XRS2, or RAD59) are inviable (Symington, 1998). Deletions may also result from defective flap processing (Figure 7.1b) if the lagging strand polymerase is stalled before meeting the 5' end of the previous Okazaki fragment. The resulting ss gap could allow secondary structure formation, which would effectively shorten the lagging template strand.

In all of these TR plasmids, the TR tracts are cloned into the EcoRI site of the multiple cloning site, which is approximately 1kb away from the unidirectional origin of plasmid replication (*rep*). Since the average size of an Okazaki fragment in *E. coli* is estimated to be between 900b and 1200b (Blumenthal and Clark, 1977) the first Okazaki fragment on the lagging strand is likely to finish within or close to the TR tract. It would therefore seem most likely that the role of DNA Polymerase I in preventing TR instability is indeed in the processing of Okazaki fragments. However, this could be verified by cloning the TR tracts closer to the plasmid origin of replication.

It should be noted that all of the models proposed for TR tract instability mediated by defective Okazaki fragment processing predict elevated rates of instability on the lagging strand of DNA synthesis. The polarity of TR tract expansions in *E. coli* (Kang *et al.,* 1995) and yeast (Maurer, O'Callaghan and Livingston, 1998) have been measured using an inserted CAT triplet in a (CAG) tract as a marker. Both studies report that $(CAG)_n$ tract expansions occur more frequently at the 5' end of the nascent lagging strand, consistent with instability effects mediated by 5' end processing. Additionally, *polA*$^-$ mutations have previously been demonstrated to destabilise dinucleotide repeat tracts in *E. coli*, partially independently of SOS induction (Morel *et al.,* 1998).

Figure 7.7: The location of Okazaki fragment termination and subsequent 5' flap processing could mediate TR tract instability



Adapted from Maurer *et al.* 1998, and Spiro *et al.* 1999.

polB⁻

DNA Polymerase II is a poorly characterised polymerase whose expression from *polB* is elevated as a consequence of SOS de-repression, but exhibits high fidelity, error-free DNA replication. Genetic studies have demonstrated that DNA Pol II may be involved in repair of UV-irradiated or oxidised DNA, as well as inter-strand cross-links (Berardini, Foster, and Loechler, 1999) and bypass of abasic lesions. DNA Pol II has recently been shown to be required for replication restart in UV-irradiated *E. coli* (Rangarajan, Woodgate, and Goodman, 1999). *polB*⁻ cells exhibit a 50 minute delay in replication restart after UV damage, but do not have reduced viability compared to UV-irradiated wild type cells, which display a 10 minute delay in replication restart. It is likely that PriA functions in the same replication restart pathway as DNA Pol II, as *priA*⁻ cells exhibit a delay of approximately 40 min in the resumption of replication. *polB*⁻ Δ*umuDC*⁻ cells exhibit delays of over 90 min, suggesting independent pathways of replication restart.

Genetic analysis of delayed replication restart epistasis groups has led to the suggestion of a rapid replication restart pathway able to act at stalled fork substrates generated in UV-irradiated cells (Rangarajan, Woodgate, and Goodman, 2000). A replication fork passing through UV-irradiated DNA encounters a lesion on one template strand, uncoupling the polymerases so that one stalls at the lesion whilst the other progresses for some distance. This generates a ss gap, which is coated with RecA (aided by RecF, RecO, and RecR). The stalled replication fork then undergoes RecA-mediated reversal, generating a chickenfoot structure with a ss overhang. The ss overhang acts as a template for DNA Polymerase II replication, which generates a long DSE chickenfoot. It would then be possible to branch migrate this structure through the DNA lesion, and provide a substrate for PriA-mediated priming of DNA Polymerase III for replication restart on the other side of the lesion.

DNA Pol II is expressed at approximately 50 molecules per cell before SOS induction, and is an early response gene in this operon due to its weak LexA binding site, so may mediate some replication restart events in cells not induced for the SOS response. However, it is unclear whether secondary structures in TR tracts uncouple replication forks in the manner suggested to provide a substrate for this pathway. The

levels of $(CCG/CGG)_{24}$ tract instability are very similar in *polB*⁻ cells, and are significantly less stable than in wild type cells. The $(CAG/CTG)_{43}$ repeats show a strong orientation-dependence on TR tract instability in *polB*⁻ cells, with $(CTG)_{43}$ on the lagging strand displaying more than twice the instability of $(CAG)_{43}$. This is consistent with DNA Polymerase II having some responsibility for restarting DNA replication after polymerase uncoupling in long TR tracts by hairpin formation on the lagging strand.

holD⁻

The *holD* gene encodes the ψ subunit of the DNA polymerase III holoenzyme, a component of the γ complex clamp loader. Therefore *holD*⁻ mutants have reduced polymerase processivity due to defective clamp loading (particularly on the lagging strand). The *holD*⁻ strain displayed elevated rates of TR tract deletion for all TR plasmids tested. This result would suggest that reducing the processivity of replication encourages forks to pause, and in TR tracts this leads to elevated rates of deletion products. It is possible that the frequent stalling of replication forks thought to occur in *holD*⁻ mutants would promote secondary structure formation because of the greater longevity of ssDNA between incomplete Okazaki fragments on the lagging strand. This may in turn promote TR deletions mediated by strand slippage, SbcCD processing mechanisms, or recombinational restart of the stalled forks. Indeed, a *holD*⁻ strain has previously been shown to increase the deletion frequency of tandem repeats (Flores *et al.,* 2001). These deletion events were demonstrated to require RecA and RecBCD, suggesting that they are a result of unequal crossing-over. It is thought that in *holD*⁻ cells, recombination is used to re-initiate reversed replication forks using the annealed nascent strands to re-invade the chromosomal duplex (see Figure 1.11 in Introduction chapter1). This is the same mechanism suggested to occur during stalled fork processing previously observed in *dnaBts* cells (Saveson and Lovett, 1997), which similarly displays tandem repeat deletions and a dependence on RecA and RecBCD (Michel, Ehrlich, and Uzest, 1997), and also cause RuvABC-dependent DSBs (Seigneur *et al.,* 1998).

UV excision repair

None of the UV excision repair mutants tested with CCG/CGG_{24} tracts had any significant effect on TR instability compared to wild type. This is in contrast to observations using $(CAG/CTG)_{175}$ on ColE1 type plasmids in *E. coli* when *uvrA*⁻ cells displayed elevated instability and *uvrB*⁻ cells displayed decreased instability compared to wild type cells (Parnewski, *et al.*, 1999). These effects varied with repeat orientation regarding the direction of replication and transcription. However, identical experiments with $(CAG/CTG)_{50}$ failed to find any differences in TR tract instability between wild type in a variety of nucleotide excision repair (NER) deficient strains, suggesting that any effects of this DNA repair pathway are only relevant once TR tracts have expanded beyond at least 50 triplets. Duplicate experiments with (CCG/CGG) tracts up to 59 triplets in length again found no difference in TR tract instability between wild type and NER-deficient cells, and the work presented in this chapter corroborates this. NER is believed to be initiated by UvrA binding to a variety of DNA substrates, including DNA cross-links, bulky DNA adducts distorting the duplex, and significantly, DNA loops. *In vitro* studies using synthetic slipped strand intermediate substrates demonstrated UvrABC binding and exonuclease action. Recent work by Oussatcheva and co-workers (2001) has shown that $(CAG/CTG)_{79}$ tract deletions are reduced in *uvrA*⁻ strains, and that transformation of DNA containing loops of 23 (CTG) triplets were less efficiently excised in *uvrA*⁻cells. *In vitro*, purified UvrA was found to bind DNA loops with approximately 100 times greater affinity than linear duplex. Clearly loops of this magnitude could only occur in long TR tracts, and as such, may reflect one mechanism of anticipation in TRD pedigrees, rather than the initial development of pre-mutation tract lengths.

Conclusions

It seems probable that TR tracts are able to change length via a range of DNA processing pathways in which the exposure of ssDNA allows secondary structure formation. It is therefore not surprising that previous studies have detected modified patterns of TR instability due to less processive replication, defective Okazaki fragment processing, mismatch repair, UV excision repair, and recombination. In the experiments performed in this chapter, factors were identified which elevated TR tract deletion rates. They included replication fork processivity (*holD*$^-$) and Okazaki fragment processing (*polA*$^-$). In addition, SOS induction (*lexA3,51 (def)*) significantly destabilised $(CAG/CTG)_{43}$ tracts. The possible activation of a DNA damage checkpoint (*umuD*$^-$) may be important in stabilising $(CAG/CTG)_{43}$ tracts in *E. coli*. Defective mismatch repair was seen to result in the distinctive and well-characterised phenomenon of +1/-1 triplet changes in TR tract length, reducing the overall stability of $(CCG/CGG)_{24}$ tracts.

TR tract length and orientation were thought to be crucial in determining tract instability, and these observations support the hypothesis that hairpin formation is a crucial factor in dynamic mutations of tract length. Studies in chapter 4 of this thesis that identified TR tract-stimulated plasmid dimerisation in certain recombination-deficient mutants, did not correlate with TR tract instability in those same mutants as measured by these end-labelling experiments. This suggests that recombination in $(CAG/CTG)_{43}$ tracts is a faithful process which does not normally lead to dynamic mutation. In addition, high deletion rates for $(CTG)_{43}$ tracts were observed in *recA*$^-$ cells, suggesting that the pattern of instability observed most frequently in these experiments was independent of homologous recombination. The short, incremental changes in TR tract length observed are consistent with replication slippage. However, the studies in this chapter did not offer the potential for gene conversion from a longer TR tract, nor was DSBR assessed satisfactorily.

Large expansion products were not detected in these experiments, perhaps as a consequence of experimental design. It may be that the TR tracts were not present for long enough in each strain for a TR phenotype to be easily observed, since deletions have previously been reported to occur far more frequently than expansions in *E. coli*.

Also, the longest TR tracts used in this study contained 43 triplets, which is far shorter than those used in other studies. The concept of using TR tracts of equivalent length to the "pre-mutation" TRD tracts in humans was intended to duplicate the mechanisms of TR tract instability acting at these repeat lengths. However, the association between TR tract length and instability has been clearly demonstrated in numerous other studies, and a better experimental approach may indeed be to identify mechanisms of instability under conditions which optimise hairpin stability (i.e. longer TR tracts). In addition there are certain differences between prokaryotic and eukaryotic DNA metabolism which may mediate TR tract instability in humans, but are not directly measurable in *E. coli*, such as meiotic recombination. In studies of human parent-to-child transmission of TR tracts, expansions have been reported to occur between three and over a hundred times more frequently than deletions (Ashizawa *et al.*, 1994; Leeflang *et al.*, 1995; McMurry, 1995). These results may reflect a general and inherent bias of human DNA metabolism at TR tracts (possibly during meiosis), or may simply indicate a bias within the sub-populations of TRD pedigrees examined.

Okazaki fragment processing by RAD27 (equivalent to human *FEN1*) has been observed to de-stabilize TR tracts in yeast (Freudenreich, Kantrow and Zakian, 1998). This is an especially significant study because one of the findings was that the instability was biased towards expansion products. PCR and Southern-hybridisation to determine TR tract length revealed that in a *rad27*$^-$ strain, $(CTG)_{40}$ tracts underwent 36-fold more frequent expansions and 10-fold more frequent deletions, and $(CTG)_{70}$ tracts underwent 33-fold more frequent expansions and 8-fold more frequent deletions than when present in a wild type background. However, it has been suggested that Okazaki fragments are far longer in *E. coli* (900-1200b according to Blumenthal and Clark, 1977) than in eukaryotes (100-200b according to Okazaki *et al.*, 1968; Wu *et al.*, 1992; Sarkar *et al.*, 1998), meaning that examination of this hypothesis would require tremendously long TR tracts in *E. coli*. Sarkar and co-workers (1998) have performed studies in *E. coli* using exceptionally long TR tracts. Only small changes in initial (CTG) tract length below 120 to 200 triplets were observed, whilst very large amplifications from 2 to 6kb were frequently found when using (CTG) tracts initially measuring 330 to 500 triplets. A twelve-fold increase in the recovery of expansion products was found when $(CTG)_{500}$ tracts were tested, when compared to $(CTG)_{120}$

tracts. It is the change in instability pattern, from frequent short deletions, to frequent large expansions, which is significant here, as these results cannot be explained by simply extrapolating the observation that general instability is proportional to tract length. This work clearly offers a means of distinguishing between two mechanisms of TR tract instability, and although it was used to demonstrate that large expansions required the absence of SbcC (so required stable hairpin formation), any effects of disrupted Okazaki flap processing (such as in a *polA*$^-$ strain) were not investigated. This is unfortunate, as it may prove to be an experiment that determines much future research into this fascinating field.

Chapter 8: Concluding remarks

The great value of the *E. coli* system is the wide availability and diversity of very well defined mutants in a series of DNA metabolic functions. The recent development of λ *red/gam* PCR cloning was found to be a useful addition to this system, enabling rapid and precise removal of three target genes. UV survival studies of *yraN*⁻ strains suggested that although the gene has sequence homology to a nuclease superfamily containing resolvases, YraN does not have a role in Holliday junction resolution. Cell viability, plasmid multimerisation, and trinucleotide repeat instability in *yraN*⁻ cells was found to be indistinguishable from wild type, and P1 transduction of *yraN*⁻ into a variety of recombination-deficient strains was not problematic. However, the high degree of sequence conservation, and distribution of homologues suggest that YraN may have an important role in eubacteria. *In vitro* work in MaLcolm White's laboratory is pursuing the possibility that YraN is a conserved nuclease involved in other pathways of DNA repair.

To examine whether $(CAG/CTG)_{43}$ tracts are recombination hot-spots in *E. coli*, plasmid dimerisation was used as an assay for crossover production during recombination. Recent work has suggested biased Holliday junction resolution for different recombination substrates. Recombinational repair of ss gaps, leading strand breaks and re-initiation of reversed forks by recombination initiated at annealed nascent strand ends would all be predicted to produce non-crossover products, so would not be detected by this assay. However, ends-in DSBR and lagging strand breaks would be predicted to produce crossovers and be detected. In this way, the similar dimerisation values in wild type cells displayed by pUC18, CAG_{43}, and CTG_{43} plasmids would either suggest that any recombination occurring in these TR tracts is initiated at substrates biased against crossover production, or that recombination at substrates formed in these TR tracts biased towards crossover production occurs so infrequently that they cannot be detected against the background level of dimerisation in pUC18. An increased recovery of dimeric CTG_{43} plasmid compared to CAG_{43} and pUC18 was found in *recF*⁻, *recQ*⁻ and *sbcC*⁻ cells. The *recF*⁻ result was not simply due to a deficiency in ss gap repair, because *recO*⁻ and *recR*⁻ cells gave similar rates of dimerisation for all three plasmids. However, *recF*⁻*recO*⁻ and *recF*⁻*recR*⁻ double mutants did not display the $(CTG)_{43}$ tract-stimulated plasmid dimerisation observed in

the *recF*⁻ single mutant. This was interpreted as the production in $(CTG)_{43}$ tracts of a recombination substrate requiring RecOR, but not RecF, such as a long ss overhang. Such a substrate could be produced by hairpin cleavage in $(CTG)_{43}$ tracts on the lagging strand occurring next to ss gaps formed by incomplete synthesis of Okazaki fragments. Whilst this result is consistent with both ss gap formation due to hairpin-mediated inhibition of DNA synthesis on the lagging strand of replication, and the formation of DSBs in $(CTG)_{43}$ tracts due to hairpin cleavage, it is probably not the dominant mechanism of instability in $(CTG)_{43}$ tracts, as end-labelling experiments of these tracts in *recF*⁻ cells detected only wild type levels of instability. Other recombination mutants that displayed elevated plasmid dimerisation (dependent on recombination) due to the presence of TR tracts, do not give elevated TR tract instability as measured by DNA end labelling.

The persistence of ss overhangs, ss gaps, DSBs, stalled and broken replication forks was investigated using *sfiA-lac* and *λ-gal* assays. Both assays demonstrated that the SOS response was not de-repressed by the presence of $(CAG/CTG)_{43}$, or $(CCG/CGG)_{24}$ tracts in *E. coli*. It would therefore seem that recombination substrates formed in TR tracts on high copy number plasmids are processed quickly enough to prevent their accumulation in the cell and induction of the SOS response. The mechanism of TR tract instability therefore seems unlikely to involve saturation of the host cell's DNA repair capability (including recombinational repair).

However, differences in DNA maintenance between TR tract and standard plasmid DNA were detected after 12 day incubations following co-transformation with copies of pUC18 and pUC18 carrying TR tracts. Plasmids carrying short TR tracts of $(CCG/CGG)_{24}$, $(CAG/CTG)_{25}$, and $(CAG/CTG)_{28}$ displayed a surprising propensity to out-compete standard pUC18, interpreted as an ability of these tracts to initiate replication. The substrates from which replication could be initiated are not clear, but suggestions presented here include: stable R-loop structures formed by transcription stalled within TR tracts, D-loops formed during DSBR of cleaved hairpins, or translesion synthesis during the SOS response. Future incubations in which the co-transformants are maintained in exponential phase growth could be helpful in distinguishing between these mechanisms. In sharp contrast, plasmids carrying longer

$(CAG/CTG)_{43}$ tracts displayed stark losses to standard pUC18, perhaps due to replication inhibition or elevated frequencies of DSB formation.

Radioactive end-labelling of excised $(CCG/CGG)_{24}$ and $(CAG/CTG)_{43}$ tracts was used to examine tract instability in a variety of DNA replication, recombination, and repair mutants. All strains exhibited a greater propensity for deletion than expansion in all tracts. This phenomenon has also been reported by others using the *E. coli* model system, and contrasts with work performed in yeast, which seems to favour expansion of TR tracts rather than deletion. It is unclear whether this is a consequence of shorter Okazaki fragments in eukaryotes, or differences in DNA repair. In the work presented here, small changes in tract length were far more common than larger changes. In general, the longer $(CAG/CTG)_{43}$ tracts appeared to be more unstable than the shorter $(CCG/CGG)_{24}$ tracts, and in both cases the orientation which displayed the greatest instability was the one where the triplet sequence thought to form the most stable secondary structures were present on the lagging strand. Both of these results support the theory that secondary structure formation is intrinsically linked to TR tract instability.

TR tracts were still unstable in *recA*$^-$ cells, but $(CAG/CTG)_{43}$ tracts were slightly more stable than in wild type cells. This would suggest that recombination is not the sole cause of instability in TR tracts of this length, and whilst other work presented here suggests that it may have a role in TR tract maintenance, it seems on the whole to remain a faithful process. Mutants found to destabilise TR tracts included *umuD*$^-$ (implicated in translesion synthesis and DNA checkpoint activities), the exonuclease activity of PolA (implicated in Okazaki fragment processing), and *holD*$^-$ (required for full processivity of replication forks). Thus it would appear that TR tract instability can occur in any aspect of DNA metabolism in which the tract is present as a single strand. In addition, loss of SbcCD hairpin endonuclease specifically increased the instability of $(CTG)_{43}$ tracts present on the lagging strand, predominantly by elevating the frequency of tract expansions. This may reflect a role for SbcCD in hairpin cleavage within TR tracts.

Appendix: Mouse flanking sequences

GAATTCCTCC CCTTTCTAGC CTTCTTCAAG CATCTTGGGA GCATCTTTGC
TGCTGCTGCT GCTGCTGCTG CTGCTGCTGC TGCTGCTGCT GCTGCTGCTG
CTGCTGCTGC TGCTGCTGCT GCTGAGATGA TCAGCAGCAG GCTCACTGCT
CAGCATCCCG TTTGGACCAA ACTGAATTC

Mini-palindromes which may act as clamps to stabilise secondary structures within the mouse $(CTG/CAG)_{25}$ tract are highlighted/underlined. The mouse $(CTG/CAG)_{43}$ contains the same flanking sequences.

Bibliography

References cited:

Abbott C, Chambers D (1994) Analysis of CAG trinucleotide repeats from mouse cDNA sequences. *Ann Hum Genet* 58 (Pt 2):87-94

Abbott PJ (1985) Stimulation of recombination between homologous sequences on carcinogen-treated plasmid DNA and chromosomal DNA by induction of the SOS response in Escherichia coli K12. *Mol Gen Genet* 201:129-132

Al-Deib AA, Mahdi AA, Lloyd RG (1996) Modulation of recombination and DNA repair by the RecG and PriA helicases of Escherichia coli K-12. *J Bacteriol* 178:6782-6789

Allers T, Leach DR (1995) DNA palindromes adopt a methylation-resistant conformation that is consistent with DNA cruciform or hairpin formation in vivo. *J Mol Biol* 252:70-85

Amundsen SK, Taylor AF, Chaudhury AM, Smith GR (1986) recD: the gene for an essential third subunit of exonuclease V. *Proc Natl Acad Sci U S A* 83:5558-5562

Anderson DG, Kowalczykowski SC (1997) The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* 90:77-86

Anderson DG, Kowalczykowski SC (1997) The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev* 11:571-581

Anderson DG, Kowalczykowski SC (1998) SSB protein controls RecBCD enzyme nuclease activity during unwinding: a new role for looped intermediates. *J Mol Biol* 282:275-285

Aravind L, Makarova KS, Koonin EV (2000) SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res* 28:3417-3432

Asai T, Sommer S, Bailone A, Kogoma T (1993) Homologous recombination-dependent initiation of DNA replication from DNA damage-inducible origins in Escherichia coli. *EMBO* 12:3287-3295

Asai T, Bates DB, Kogoma T (1994) DNA replication triggered by double-stranded breaks in E. coli: dependence on homologous recombination functions. *Cell* 78:1051-1061

Asai T, Kogoma T (1994) D-loops and R-loops: alternative mechanisms for the initiation of chromosome replication in Escherichia coli. *J Bacteriol* 176:1807-1812

Ashizawa T, Anvret M, Baiget M, Barcelo JM, Brunner H, Cobo AM, Dallapiccola B, Fenwick RGJ, Grandell U, Harley H (1994) Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *Am J Hum Genet* 54:414-423

Bacolla A, Gellibolian R, Shimizu M, Amirhaeri S, Kang S, Ohshima K, Larson JE, Harvey SC, Stollar BD, Wells RD (1997) Flexible DNA: genetically unstable CTG.CAG and CGG.CCG from human hereditary neuromuscular disease genes. *J Biol Chem* 272:16783-16792

Balakumaran BS, Freudenreich CH, Zakian VA (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Hum Mol Genet* 9:93-100

Balbinder E, Coll B, Hutchinson J, Bianchi AS, Groman T, Wheeler KA, Meyer M (1993) Participation of the SOS system in producing deletions in E. coli plasmids. *Mutat Res* 286:253-265

Barre FX, Soballe B, Michel B, Aroyo M, Robertson M, Sherratt D (2001) Circles: the replication-recombination-chromosome segregation connection. *Proc Natl Acad Sci U S A* 98:8189-8195

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C (1993) A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res* 21:3329-3330

Behnke D, Malke H, Hartmann M, Walter F (1979) Post-transformational rearrangement of an in vitro reconstructed group-A streptococcal erythromycin resistance plasmid. *Plasmid* 2:605-616

Bennett RJ, Dunderdale HJ, West SC (1993) Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell* 74:1021-1031

Bennett RJ, West SC (1996) Resolution of Holliday junctions in genetic recombination: RuvC protein nicks DNA at the point of strand exchange. *Proc Natl Acad Sci U S A* 93:12217-12222

Berardini M, Foster PL, Loechler EL (1999) DNA polymerase II (polB) is involved in a new DNA repair pathway for DNA interstrand cross-links in Escherichia coli. *J Bacteriol* 181:2878-2882

Bi X, Lyu YL, Liu LF (1995) Specific stimulation of recA-independent plasmid recombination by a DNA sequence at a distance. *J Mol Biol* 247:890-902

Bi X, Liu LF (1996) A replicational model for DNA recombination between direct repeats. *J Mol Biol* 256:849-858

Bidnenko V, Seigneur M, Penel-Colin M, Bouton MF, Dusko ES, Michel B (1999) sbcB sbcC null mutations allow RecF-mediated repair of arrested replication forks in rep recBC mutants. *Mol Microbiol* 33:846-857

Bidnenko V, Ehrlich SD, Michel B (2002) Replication fork collapse at replication terminator sequences. *EMBO J* 21:3898-3907

Bierne H, Michel B (1994) When replication forks stop. *Mol Microbiol* 13:17-23

Bierne H, Seigneur M, Ehrlich SD, Michel B (1997) uvrD mutations enhance tandem repeat deletion in the Escherichia coli chromosome via SOS induction of the RecF recombination pathway. *Mol Microbiol* 26:557-567

Bierne H, Ehrlich SD, Michel B (1997) Deletions at stalled replication forks occur by two different pathways. *EMBO* 16:3332-3340

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523

Blumenthal AB, Clark EJ (1977) Discrete sizes of replication intermediates in Drosophila cells. 12:183-189

Bonner CA, Hays S, McEntee K, Goodman MF (1990) DNA polymerase II is encoded by the DNA damage-inducible dinA gene of Escherichia coli. *Proc Natl Acad Sci U S A* 87:7663-7667

Bork JM, Cox MM, Inman RB (2001) The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA. *EMBO* 20:7313-7322

Bork JM, Cox MM, Inman RB (2001) RecA protein filaments disassemble in the 5' to 3' direction on single-stranded DNA. *J Biol Chem* 276:45740-45743

Boucher CA, King SK, Carey N, Krahe R, Winchester CL, Rahman S, Creavin T, Meghji P, Bailey ME, Chartier FL (1995) A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable (CTG)n repeat. *Hum Mol Genet* 4:1919-1925

Bowater RP, Rosche WA, Jaworski A, Sinden RR, Wells RD (1996) Relationship between Escherichia coli growth and deletions of CTG.CAG triplet repeats in plasmids. *J Mol Biol* 264:82-96

Bowater RP, Jaworski A, Larson JE, Parniewski P, Wells RD (1997) Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli. *Nucleic Acids Res* 25:2861-2868

Bowater RP, Wells RD (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog Nucleic Acid Res Mol Biol* 66:159-202

Bravo A, Alonso JC (1990) The generation of concatemeric plasmid DNA in Bacillus subtilis as a consequence of bacteriophage SPP1 infection. *Nucleic Acids Res* 18:4651-4657

Brent R, Ptashne M (1980) The lexA gene product represses its own promoter. *Proc Natl Acad Sci U S A* 77:1932-1936

Bridges BA (1998) DNA repair: getting past a lesion - at a cost. *Curr Biol* 8:R886-R888

Burckhardt SE, Woodgate R, Scheuermann RH, Echols H (1988) UmuD mutagenesis protein of Escherichia coli: overproduction, purification, and cleavage by RecA. *Proc Natl Acad Sci U S A* 85:1811-1815

Bzymek M, Lovett ST (2001) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* 98:8319-8325

Bzymek M, Lovett ST (2001) Evidence for two mechanisms of palindrome-stimulated deletion in Escherichia coli: single-strand annealing and replication slipped mispairing. *Genetics* 158:527-540

Chalker AF, Leach DR, Lloyd RG (1988) Escherichia coli sbcC mutants permit stable propagation of DNA replicons containing a long palindrome.Gene 71(1):201-5.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci U S A* 94:1041-1046

Chakraverty RK, Hickson ID (1999) Defending genome integrity during DNA replication: a proposed role for RecQ family helicases. *Bioessays* 21:286-294

Chalker AF, Leach DR, Lloyd RG (1988) Escherichia coli sbcC mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* 71:201-205

Chambers SP, Prior SE, Barstow DA, Minton NP (1988) The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* 68:139-149

Chastain PD, Eichler EE, Kang S, Nelson DL, Levene SD, Sinden RR (1995) Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry* 34:16125-16131

Chastain PD, Sinden RR (1998) CTG repeats associated with human genetic disease are inherently flexible. *J Mol Biol* 275:405-411

Chaudhury AM, Smith GR (1984) Escherichia coli recBC deletion mutants. *J Bacteriol* 160:788-791

Chaudhury AM, Smith GR (1985) Role of Escherichia coli RecBC enzyme in SOS induction. *Mol Gen Genet* 201:525-528

Chen W, Jinks-Robertson S (1998) Mismatch repair proteins regulate heteroduplex formation during mitotic recombination in yeast. *Mol Cell Biol* 18:6525-6537

Cheng RZ, Murano S, Kurz B, Shmookler R Homologous recombination is elevated in some Werner-like syndromes but not during normal in vitro or in vivo senescence of mammalian cells. *Mutat Res* 237:259-269

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, Haddad BR, Richards F, Sillence D, Greenberg CR, Ives E, Van den EG, Hughes MR, Hayden MR (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. *Hum Mol Genet* 6:301-309

Christiansen J, Kofod M, Nielsen FC (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Res* 22:5709-5716

Chua KL, Mak YK, Oliver P (1993) Expression of the recA gene in recombination-deficient (rec-) strains of Escherichia coli. *Biochimie* 75:775-783

Clark AJ (1971) Toward a metabolic interpretation of genetic recombination of E. coli and its phages. *Annu Rev Microbiol* 25:437-464

Clark AJ (1991) rec genes and homologous recombination proteins in Escherichia coli. *Biochimie* 73:523-532

Clark HB, Orr HT (2000) Spinocerebellar ataxia type 1--modeling the pathogenesis of a polyglutamine neurodegenerative disorder in transgenic mice. *J Neuropathol Exp Neurol* 59:265-270

Clegg RM, Murchie AI, Lilley DM (1994) The solution structure of the four-way DNA junction at low-salt conditions: a fluorescence resonance energy transfer analysis. *Biophys J* 66:99-109

Clewell DB (1972) Nature of Col E 1 plasmid replication in Escherichia coli in the presence of the chloramphenicol. *J Bacteriol* 110:667-676

Cobbe N, Heck MM (2000) Review: SMCs in the world of chromosome biology- from prokaryotes to higher eukaryotes. *J Struct Biol* 129:123-143

Cohen A, Laban A (1983) Plasmidic recombination in Escherichia coli K-12: the role of recF gene function. *Mol Gen Genet* 189:471-474

Cohen A, Silberstein Z, Broido S, Laban A (1985) General genetic recombination of bacterial plasmids. *Basic Life Sci* 30:505-519

Cohen A, Clark AJ (1986) Synthesis of linear plasmid multimers in Escherichia coli K-12. *J Bacteriol* 167:327-335

Cohen A, Clark AJ (1986) Synthesis of linear plasmid multimers in Escherichia coli K-12. *J Bacteriol* 167:327-335

Collister M, Lane DP, Kuehl BL (1998) Differential expression of p53, p21waf1/cip1 and hdm2 dependent on DNA damage in Bloom's syndrome fibroblasts. *Carcinogenesis* 19:2115-2120

Connelly JC, Leach DR (1996) The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* 1:285-291

Connelly JC, de Leau ES, Okely EA, Leach DR (1997) Overexpression, purification, and characterization of the SbcCD protein from Escherichia coli. *J Biol Chem* 272:19819-19826

Connelly JC, Kirkham LA, Leach DR (1998) The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci U S A* 95:7969-7974

Connelly JC, de Leau ES, Leach DR (1999) DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. *Nucleic Acids Res* 27:1039-1046

Connolly B, Parsons CA, Benson FE, Dunderdale HJ, Sharples GJ, Lloyd RG, West SC (1991) Resolution of Holliday junctions in vitro requires the Escherichia coli ruvC gene product. *Proc Natl Acad Sci U S A* 88:6063-6067

Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, Hickson ID, West SC (2000) Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO* 1:80-84

Constantinou A, Davies AA, West SC (2001) Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell* 104:259-268

Corrette-Bennett S.E., Lovett ST (1995) Enhancement of RecA strand-transfer activity by the RecJ exonuclease of Escherichia coli. *J Biol Chem* 270:6881-6885

Courcelle J, Carswell-Crumpton C, Hanawalt PC (1997) recF and recR are required for the resumption of replication at DNA replication forks in Escherichia coli. *Proc Natl Acad Sci U S A* 94:3714-3719

Courcelle J, Hanawalt PC (1999) RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated Escherichia coli. *Mol Gen Genet* 262:543-551

Courcelle J, Crowley DJ, Hanawalt PC (1999) Recovery of DNA replication in UV-irradiated Escherichia coli requires both excision repair and recF protein function. *J Bacteriol* 181:916-922

Courcelle J, Hanawalt PC (2001) Participation of recombination proteins in rescue of arrested replication forks in UV-irradiated Escherichia coli need not involve recombination. *Proc Natl Acad Sci U S A* 98:8196-8202

Cox MM, Lehman IR (1981) Directionality and polarity in recA protein-promoted branch migration. *Proc Natl Acad Sci U S A* 78:6018-6022

Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ, Marians KJ (2000) The importance of repairing stalled replication forks. *Nature* 404:37-41

Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ, Marians KJ (2000) The importance of repairing stalled replication forks. *Nature* 404:37-41

Cox MM (2001) Historical overview: searching for replication help in all of the rec places. *Proc Natl Acad Sci U S A* 98:8173-8180

Cox MM (2002) The nonmutagenic repair of broken replication forks via recombination. *Mutat Res* 510:107-120

Cromie GA, Leach DR (2000) Control of crossing over. *Mol Cell* 6:815-826

Cromie GA, Millar CB, Schmidt KH, Leach DR (2000) Palindromes as substrates for multiple pathways of recombination in Escherichia coli. *Genetics* 154:513-522

Cromie GA, Connelly JC, Leach DR (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell* 8:1163-1174

Cummings CJ, Mancini MA, Antalffy B, DeFranco DB, Orr HT, Zoghbi HY (1998) Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet* 19:148-154

Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* 9:909-916

Cunniff C (2001) Molecular mechanisms in neurologic disorders. *Semin Pediatr Neurol* 8:128-134

Cunningham RP, DasGupta C, Shibata T, Radding CM (1980) Homologous pairing in genetic recombination: recA protein makes joint molecules of gapped circular DNA and closed circular DNA. *Cell* 20:223-235

Dabert P, Ehrlich SD, Gruss A (1992) High-molecular-weight linear multimer formation by single-stranded DNA plasmids in Escherichia coli. *J Bacteriol* 174:173-178

Dabert P, Smith GR (1997) Gene replacement with linear DNA fragments in wild-type Escherichia coli: enhancement by Chi sites. *Genetics* 145:877-889

Dallas JF (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm Genome* 3:452-456

Dambly-Chaudiere C, Gottesman M, Debouck C, Adhya S (1983) Regulation of the pR operon of bacteriophage lambda. *J Mol Appl Genet* 2:45-56

Darlow JM, Leach DR (1995) The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in Escherichia coli suggest hairpin folding preferences in vivo. *Genetics* 141:825-832

Darlow JM, Leach DR (1998) Evidence for two preferred hairpin folding patterns in d(CGG).d(CCG) repeat tracts in vivo. *J Mol Biol* 275:17-23

Darlow JM, Leach DR (1998) Secondary structures in d(CGG) and d(CCG) repeat tracts. *J Mol Biol* 275:3-16

DasGupta C, Shibata T, Cunningham RP, Radding CM (1980) The topology of homologous pairing promoted by RecA protein. *Cell* 22:437-446

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97:6640-6645

Davey S, Han CS, Ramer SA, Klassen JC, Jacobson A, Eisenberger A, Hopkins KM, Lieberman HB, Freyer GA (1998) Fission yeast rad12+ regulates cell cycle checkpoint control and is homologous to the Bloom's syndrome disease gene. *Mol Cell Biol* 18:2721-2728

Davies AA, West SC (1998) Formation of RuvABC-Holliday junction complexes in vitro. *Curr Biol* 8:725-727

Davison A, Leach DR (1994) The effects of nucleotide sequence changes on DNA secondary structure formation in Escherichia coli are consistent with cruciform extrusion in vivo. *Genetics* 137:361-368

Dayn A, Malkhosyan S, Duzhy D, Lyamichev V, Panchenko Y, Mirkin S (1991) Formation of (dA-dT)n cruciforms in Escherichia coli cells under different environmental conditions. *J Bacteriol* 173:2658-2664

de AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, Woodgate R (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol* 35:1560-1572

Di Capua E, Engel A, Stasiak A, Koller T (1982) Characterization of complexes between recA protein and duplex DNA by electron microscopy. *J Mol Biol* 157:87-103

Dickman MJ, Ingleston SM, Sedelnikova SE, Rafferty JB, Lloyd RG, Grasby JA, Hornby DP (2002) The RuvABC resolvasome. *Eur J Biochem* 269:5492-5501

Dixon DA, Kowalczykowski SC (1993) The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell* 73:87-96

Doe CL, Dixon J, Osman F, Whitby MC (2000) Partial suppression of the fission yeast rqh1(-) phenotype by expression of a bacterial Holliday junction resolvase. *EMBO J* 19:2751-2762

Dooley TP, Polisky B (1987) Suppression of ColE1 RNA-RNA mismatch mutations in vivo by the ColE1 Rop protein. *Plasmid* 18:24-34

Dower NA, Stahl FW (1981) Chi activity during transduction-associated recombination. *Proc Natl Acad Sci U S A* 78:7033-7037

Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* 16:6127-6145

Dunderdale HJ, Benson FE, Parsons CA, Sharples GJ, Lloyd RG, West SC (1991) Formation and resolution of recombination intermediates by E. coli RecA and RuvC proteins. *Nature* 354:506-510

Dunman PM, Ren L, Rahman MS, Palejwala VA, Murphy HS, Volkert MR, Humayun MZ (2000) Escherichia coli cells defective for the recN gene display constitutive elevation of mutagenesis at 3,N(4)-ethenocytosine via an SOS-induced mechanism. *Mol Microbiol* 37:680-686

Dutreix M, Rao BJ, Radding CM (1991) The effects on strand exchange of 5' versus 3' ends of single-stranded DNA in RecA nucleoprotein filaments. *J Mol Biol* 219:645-654

Eggleston AK, Mitchell AH, West SC (1997) In vitro reconstitution of the late steps of genetic recombination in E. coli. *Cell* 89:607-617

Eggleston AK, West SC (2000) Cleavage of holliday junctions by the Escherichia coli RuvABC complex. *J Biol Chem* 275:26467-26476

Eguchi Y, Ogawa T, Ogawa H (1988) Stimulation of RecA-mediated cleavage of phage phi 80 cI repressor by deoxydinucleotides. *J Mol Biol* 204:69-77

Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat Genet* 8:88-94

Elborough KM, West SC (1990) Resolution of synthetic Holliday junctions in DNA by an endonuclease activity from calf thymus. *EMBO* 9:2931-2936

Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, Proytcheva M, German J (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83:655-666

Ellis NA (1996) Mutation-causing mutations. *Nature* 381:110-111

Ellis NA, German J (1996) Molecular genetics of Bloom's syndrome. *Hum Mol Genet* 5 Spec No:1457-1463

Evert BO, Wullner U, Klockgether T (2000) Cell death in polyglutamine diseases. *Cell Tissue Res* 301:189-204

Feng Y, Gutekunst CA, Eberhart DE, Yi H, Warren ST, Hersch SM (1997) Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J Neurosci* 17:1539-1547

Fernandez de Henestrosa, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, Woodgate R (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol* 35:1560-1572

Finch PW, Chambers P, Emmerson PT (1985) Identification of the Escherichia coli recN gene product as a major SOS protein. *J Bacteriol* 164:653-658

Fishel RA, James AA, Kolodner R (1981) recA-independent general genetic recombination of plasmids. *Nature* 294:184-186

Flores MJ, Bierne H, Ehrlich SD, Michel B (2001) Impairment of lagging strand synthesis triggers the formation of a RuvABC substrate at replication forks. *EMBO* 20:619-629

Flores MJ, Ehrlich SD, Michel B (2002) Primosome assembly requirement for replication restart in the Escherichia coli holDG10 replication mutant. *Mol Microbiol* 44:783-792

Flory J, Tsang SS, Muniyappa K (1984) Isolation and visualization of active presynaptic filaments of recA protein and single-stranded DNA. *Proc Natl Acad Sci U S A* 81:7026-7030

Fouts KE, Wasie-Gilbert T, Willis DK, Clark AJ, Barbour SD (1983) Genetic analysis of transposon-induced mutations of the Rac prophage in Escherichia coli K-12 which affect expression and function of recE. *J Bacteriol* 156:718-726

Frei C, Gasser SM (2000) RecQ-like helicases: the DNA replication checkpoint connection. *J Cell Sci* 113 (Pt 15):2641-2646

Frei C, Gasser SM (2000) The yeast Sgs1p helicase acts upstream of Rad53p in the DNA replication checkpoint and colocalizes with Rad53p in S-phase-specific foci. *Genes Dev* 14:81-96

Freudenreich CH, Stavenhagen JB, Zakian VA (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol* 17:2090-2098

Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279:853-856

Friedman-Ohana R, Cohen A (1998) Heteroduplex joint formation in Escherichia coli recombination is initiated by pairing of a 3'-ending strand. *Proc Natl Acad Sci U S A* 95:6909-6914

Friedman-Ohana R, Karunker I, Cohen A (1999) A RecG-independent nonconservative branch migration mechanism in Escherichia coli recombination. *J Bacteriol* 181:7199-7205

Fry M, Loeb LA (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A* 91:4950-4954

Fry M, Loeb LA (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J Biol Chem* 274:12797-12802

Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* 81:533-540

Gacy AM, McMurray CT (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry* 37:9426-9434

Gangloff S, Soustelle C, Fabre F (2000) Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nat Genet* 25:192-194

Gellert M, O'Dea MH, Mizuuchi K (1983) Slow cruciform transitions in palindromic DNA. *Proc Natl Acad Sci U S A* 80:5545-5549

Gellibolian R, Bacolla A, Wells RD (1997) Triplet repeat instability and DNA topology: an expansion model based on statistical mechanics. *J Biol Chem* 272:16793-16797

George JW, Kreuzer KN (1996) Repair of double-strand breaks in bacteriophage T4 by a mechanism that involves extensive DNA replication. *Genetics* 143:1507-1520

Goldberg YP, McMurray CT, Zeisler J, Almqvist E, Sillence D, Richards F, Gacy AM, Buchanan J, Telenius H, Hayden MR (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. *Hum Mol Genet* 4:1911-1918

Goldberg YP, Nicholson DW, Rasper DM, KaLchman MA, Koide HB, Graham RK, Bromm M, Kazemi-Esfarjani P, Thornberry NA, Vaillancourt JP, Hayden MR (1996) Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nat Genet* 13:442-449

Gonda DK, Radding CM (1986) The mechanism of the search for homology promoted by recA protein. Facilitated diffusion within nucleoprotein networks. *J Biol Chem* 261:13087-13096

Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 71:17-50

Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol Cell Biol* 13:5315-5322

Gordenin DA, Kunkel TA, Resnick MA (1997) Repeat expansion--all in a flap? *Nat Genet* 16:116-118

Gough GW, Sullivan KM, Lilley DM (1986) The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite. *EMBO* 5:191-196

Grabczyk E, Usdin K (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res* 28:2815-2822

Grabczyk E, Kumari D, Usdin K (2001) Fragile X syndrome and Friedreich's ataxia: two different paradigms for repeat induced transcript insufficiency. *Brain Res Bull* 56:367-373

Gray MD, Shen JC, A.S., Blank A, Sopher BL, Martin GM, Oshima J, Loeb LA (1997) The Werner syndrome protein is a DNA helicase. *Nat Genet* 17:100-103

Griffith JD, Harris LD, Register J3 (1984) Visualization of SSB-ssDNA complexes active in the assembly of stable RecA-DNA filaments. *Cold Spring Harb Symp Quant Biol* 49:553-559

Grompone G, Seigneur M, Ehrlich SD, Michel B (2002) Replication fork reversal in DNA polymerase III mutants of Escherichia coli: a role for the beta clamp. *Mol Microbiol* 44:1331-1339

Gueron M, Kochoyan M, Leroy JL (1987) A single mode of DNA base-pair opening drives imino proton exchange. *Nature* 328:89-92

Gunter C, Paradee W, Crawford DC, Meadows KA, Newman J, Kunst CB, Nelson DL, Schwartz C, Murray A, Macpherson JN, Sherman SL, Warren ST (1998) Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum Mol Genet* 7:1935-1946

Gurney ME (2000) What transgenic mice tell us about neurodegenerative disease. *Bioessays* 22:297-304

Haber JE (1999) DNA recombination: the replication connection. *Trends Biochem Sci* 24:271-275

Hall SD, Kolodner RD (1994) Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. *Proc Natl Acad Sci U S A* 91:3205-3209

Han J, Hsu C, Zhu Z, Longshore JW, Finley WH (1994) Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. *Nucleic Acids Res* 22:1735-1740

Hanada K, Ukita T, Kohno Y, Saito K, Kato J, Ikeda H (1997) RecQ DNA helicase is a suppressor of illegitimate recombination in Escherichia coli. *Proc Natl Acad Sci U S A* 94:3860-3865

Harmon FG, Kowalczykowski SC (1998) RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev* 12:1134-1144

Harmon FG, DiGate RJ, Kowalczykowski SC (1999) RecQ helicase and topoisomerase III comprise a novel DNA strand passage function: a conserved mechanism for control of DNA recombination. *Mol Cell* 3:611-620

Hargreaves D, Rice DW, Sedelnikova SE, Artymiuk PJ, Lloyd RG, Rafferty JB (1998) Crystal structure of E.coli RuvA with bound DNA Holliday junction at 6 A resolution. *Nat Struct Biol* 5:441-446

Hargreaves D, Rafferty JB, Sedelnikova SE, Lloyd RG, Rice DW (1999) Crystallization of Escherichia coli RuvA complexed with a synthetic Holliday junction. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 1):263-265

Harmon FG, Kowalczykowski SC (1998) RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev* 12:1134-1144

Harmon FG, Kowalczykowski SC (2001) Biochemical characterization of the DNA helicase activity of the escherichia coli RecQ helicase. *J Biol Chem* 276:232-243

Harrington JJ, Lieber MR (1994) The characterization of a mammalian DNA structure-specific endonuclease. *EMBO* 13:1235-1246

Hashem VI, Klysik EA, Rosche WA, Sinden RR (2002) Instability of repeated DNAs during transformation in Escherichia coli. *Mutat Res* 502:39-46

Hegde S, Sandler SJ, Clark AJ, Madiraju MV (1995) recO and recR mutations delay induction of the SOS response in Escherichia coli. *Mol Gen Genet* 246:254-258

Helmer-Citterich M, Anceschi MM, Banner DW, Cesareni G (1988) Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. *EMBO* 7:557-566

Henderson ST, Petes TD (1993) Instability of a plasmid-borne inverted repeat in Saccharomyces cerevisiae. *Genetics* 134:57-62

Henricksen LA, Tom S, Liu Y, Bambara RA (2000) Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. *J Biol Chem* 275:16420-16427

Hertman I, Luria SE (1967) Transduction studies on the role of a rec+ gene in the ultraviolet induction of prophage lambda. *J Mol Biol* 23:117-133

Hickson ID, Davies SL, Li JL, Levitt NC, Mohaghegh P, North PS, Wu L (2001) Role of the Bloom's syndrome helicase in maintenance of genome stability. *Biochem Soc Trans* 29:201-204

Higgins NP, Kato K, Strauss B (1976) A model for replication repair in mammalian cells. *J Mol Biol* 101:417-425

Hirota Y, Yasuda S, Yamada M, Nishimura A, Sugimoto K, Sugisaki H, Oka A, Takanami M (1979) Structural and functional properties of the Escherichia coli origin of DNA replication. *Cold Spring Harb Symp Quant Biol* 43 Pt 1:129-138

Hirst MC, White PJ (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. *Nucleic Acids Res* 26:2353-2358

Hobom G, Hogness DS (1974) The role of recombination in the formation of circular oligomers of the lambda plasmid. *J Mol Biol* 88:65-87

Holliday R (1964) A mechannism for gene conversion in fungi. *Genet res* 5:282-304

Hong X, Cadwell GW, Kogoma T (1995) Escherichia coli RecG and RecA proteins in R-loop formation. *EMBO* 14:2385-2392

Hong X, Cadwell GW, Kogoma T (1996) Activation of stable DNA replication in rapidly growing Escherichia coli at the time of entry to stationary phase. *Mol Microbiol* 21:953-961

Horiuchi T, Fujimura Y (1995) Recombinational rescue of the stalled DNA replication fork: a model based on analysis of an Escherichia coli strain with a chromosome region difficult to replicate. *J Bacteriol* 177:783-791

Howard-Flanders P, Theriot L (1966) Mutants of Escherichia coli K-12 defective in DNA repair and in genetic recombination. *Genetics* 53:1137-1150

Howe WE, Mount DW (1978) Analysis of cell division in single clones of the Escherichia coli K-12 lexA mutant. *J Bacteriol* 133:1278-1281

Huisman O, D'Ari R (1981) An inducible DNA replication-cell division coupling mechanism in E. coli. *Nature* 290:797-799

Huisman O, Jacques M, D'Ari R, Caro L (1983) Role of the sfiA-dependent cell division regulation system in Escherichia coli. *J Bacteriol* 153:1072-1074

Humayun MZ (1998) SOS and Mayday: multiple inducible mutagenic pathways in Escherichia coli. *Mol Microbiol* 30:905-910

Hyde H, Davies AA, Benson FE, West SC (1994) Resolution of recombination intermediates by a mammalian activity functionally analogous to Escherichia coli RuvC resolvase. *J Biol Chem* 269:5202-5209

Ikeda H, Tomizawa JI (1965a) Transducing fragments in generalized transduction by phage P1. I. Molecular origin of the fragments. *J Mol Biol* 14:85-109

Ikeda H, Tomizawa JI (1965b) Transducing fragments in generalized transduction by phage P1. II. Association of DNA and protein in the fragments. *J Mol Biol* 14:110-119

Ikeda H, Tomizawa JI (1965c) Transducing fragments in generalized transduction by phage P1. III. Studies with small phage particles. *J Mol Biol* 14:120-129

Irino N, Nakayama K, Nakayama H (1986) The recQ gene of Escherichia coli K12: primary structure and evidence for SOS regulation. *Mol Gen Genet* 205:298-304

Ishioka K, Fukuoh A, Iwasaki H, Nakata A, Shinagawa H (1998) Abortive recombination in Escherichia coli ruv mutants blocks chromosome partitioning. *Genes Cells* 3:209-220

Iwasaki H, Takahagi M, Shiba T, Nakata A, Shinagawa H (1991) Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO J* 10:4381-4389

Iwasaki H, Takahagi M, Shiba T, Nakata A, Shinagawa H (1991) Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO* 10:4381-4389

Iwasaki H, Takahagi M, Nakata A, Shinagawa H (1992) Escherichia coli RuvA and RuvB proteins specifically interact with Holliday junctions and promote branch migration. *Genes Dev* 6:2214-2220

Iyer RR, Wells RD (1999) Expansion and deletion of triplet repeat sequences in Escherichia coli occur on the leading strand of DNA replication. *J Biol Chem* 274:3865-3877

Iyer RR, Pluciennik A, Rosche WA, Sinden RR, Wells RD (2000) DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in Escherichia coli. *J Biol Chem* 275:2174-2184

Jain SK, Cox MM, Inman RB (1995) Occurrence of three-stranded DNA within a RecA protein filament. *J Biol Chem* 270:4943-4949

Jakupciak JP, Wells RD (1999) Genetic instabilities in (CTG.CAG) repeats occur by recombination. *J Biol Chem* 274:23468-23479

Jakupciak JP, Wells RD (2000a) Genetic instabilities of triplet repeat sequences by recombination. *IUBMB Life* 50:355-359

Jakupciak JP, Wells RD (2000b) Gene conversion (recombination) mediates expansions of CTG[middle dot]CAG repeats. *J Biol Chem* 275:40003-40013

James AA, Morrison PT, Kolodner R (1983) Isolation of genetic elements that increase frequencies of plasmid recombinants. *Nature* 303:256-259

Janion C, Sikora A, Nowosielska A, Grzesiuk E (2002) Induction of the SOS response in starved Escherichia coli. *Environ Mol Mutagen* 40:129-133

Jankowski C, Nasar F, Nag DK (2000) Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc Natl Acad Sci U S A* 97:2134-2139

Jankowski C, Nag DK (2002) Most meiotic CAG repeat tract-length alterations in yeast are SPO11 dependent. *Mol Genet Genomics* 267:64-70

Jelinsky SA, Samson LD (1999) Global response of Saccharomyces cerevisiae to an alkylating agent. *Proc Natl Acad Sci U S A* 96:1486-1491

Jin R, M.E., Novick RP (1997) Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform? *EMBO* 16:4456-4466

Jin R, Novick RP (2001) Role of the double-strand origin cruciform in pT181 replication. *Plasmid* 46:95-105

Jiricny J (1998) Eukaryotic mismatch repair: an update. *Mutation Research* 409:107-121

Kaguni JM, Kornberg A (1984) Replication initiated at the origin (oriC) of the E. coli chromosome reconstituted with purified enzymes. *Cell* 38:183-190

Kang S, Ohshima K, Shimizu M, Amirhaeri S, Wells RD (1995) Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J Biol Chem* 270:27014-27021

Kang S, Jaworski A, Ohshima K, Wells RD (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet* 10:213-218

Kang S, Ohshima K, Jaworski A, Wells RD (1996) CTG triplet repeats from the myotonic dystrophy gene are expanded in Escherichia coli distal to the replication origin as a single large event. *J Mol Biol* 258:543-547

Karow JK, Chakraverty RK, Hickson ID (1997) The Bloom's syndrome gene product is a 3'-5' DNA helicase. *J Biol Chem* 272:30611-30614

Karow JK, Wu L, Hickson ID (2000) RecQ family helicases: roles in cancer and aging. *Curr Opin Genet Dev* 10:32-38

Karow JK, Constantinou A, Li JL, West SC, Hickson ID (2000) The Bloom's syndrome gene product promotes branch migration of holliday junctions. *Proc Natl Acad Sci U S A* 97:6504-6508

Kieser T, Melton RE (1988) Plasmid pIJ699, a multi-copy positive-selection vector for Streptomyces. *Gene* 65:83-91

Kim M, Lee HS, LaForet G, McIntyre C, Martin EJ, Chang P, Kim TW, Williams M, Reddy PH, Tagle D, Boyce FM, Won L, Heller A, Aronin N, DiFiglia M (1999) Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J Neurosci* 19:964-973

Kitao S, Ohsugi I, Ichikawa K, Goto M, Furuichi Y, Shimamoto A (1998) Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes. *Genomics* 54:443-452

Klement IA, Skinner PJ, Kaytor MD, Yi H, Hersch SM, Clark HB, Zoghbi HY, Orr HT (1998) Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* 95:41-53

Klesert TR, Otten AD, Bird TD, Tapscott SJ (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat Genet* 16:402-406

Knudson AGJ (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820-823

Kochoyan M, Leroy JL, Gueron M (1987) Proton exchange and base-pair lifetimes in a deoxy-duplex containing a purine-pyrimidine step and in the duplex of inverse sequence. *J Mol Biol* 196:599-609

Kogoma T, Torrey TA, Connaughton MJ (1979) Induction of UV-resistant DNA replication in Escherichia coli: induced stable DNA replication as an SOS function. *Mol Gen Genet* 176:1-9

Kogoma T, Skarstad K, Boye E, von Meyenburg K, Steen HB (1985) RecA protein acts at the initiation of stable DNA replication in rnh mutants of Escherichia coli K-12. *J Bacteriol* 163:439-444

Kogoma T, Subia NL, von Meyenburg K (1985) Function of ribonuclease H in initiation of DNA replication in Escherichia coli K-12. *Mol Gen Genet* 200:103-109

Kogoma T (1996) Recombination by replication. *Cell* 85:625-627

Kogoma T, Cadwell GW, Barnard KG, Asai T (1996) The DNA replication priming protein, PriA, is required for homologous recombination and double-strand break repair. *J Bacteriol* 178:1258-1264

Kogoma T (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61:212-238

Kogoma T (1997) Is RecF a DNA replication protein? *Proc Natl Acad Sci U S A* 94:3483-3484

Kogoma T, Maldonado RR (1997) DNA polymerase I in constitutive stable DNA replication in Escherichia coli. *J Bacteriol* 179:2109-2115

Kokoska RJ, Stefanovic L, Tran HT, Resnick MA, Gordenin DA, Petes TD (1998) Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). *Mol Cell Biol* 18:2779-2788

Kolodner R, Fishel RA, Howard M (1985) Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in Escherichia coli. *J Bacteriol* 163:1060-1066

Kolodner R, Fishel RA, Howard M (1985) Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in Escherichia coli. *J Bacteriol* 163:1060-1066

Kolodner R, Hall SD, Luisi-DeLuca C (1994) Homologous pairing proteins encoded by the Escherichia coli recE and recT genes. *Mol Microbiol* 11:23-30

Konforti BB, Davis RW (1990) The preference for a 3' homologous end is intrinsic to RecA-promoted strand exchange. *J Biol Chem* 265:6916-6920

Kopecko DJ, Cohen SN (1975) Site specific recA--independent recombination between bacterial plasmids: involvement of palindromes at the recombinational loci. *Proc Natl Acad Sci U S A* 72:1373-1377

Koshy BT, Zoghbi HY (1997) The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis. *Brain Pathol* 7:927-942

Kovtun IV, McMurray CT (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet* 27:407-411

Kowalczykowski SC, Krupp RA (1987) Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J Mol Biol* 193:97-113

Kowalczykowski SC, Krupp RA (1987) Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J Mol Biol* 193:97-113

Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM (1994) Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev* 58:401-465

Krasilnikov AS, Panyutin IG, Samadashwily GM, Cox R, Lazurkin YS, Mirkin SM (1997) Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res* 25:1339-1346

Krasilnikova MM, Samadashwily GM, Krasilnikov AS, Mirkin SM (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO* 17:5095-5102

Krasin F, Hutchinson F (1977) Repair of DNA double-strand breaks in Escherichia coli, which requires recA function and the presence of a duplicate genome. *J Mol Biol* 116:81-98

Kuan CT, Liu SK, Tessman I (1991) Excision and transposition of Tn5 as an SOS activity in Escherichia coli. *Genetics* 128:45-57

Kuhl DP, Caskey CT (1993) Trinucleotide repeats and genome variation. *Curr Opin Genet Dev* 3:404-407

Kunst CB, Leeflang EP, Iber JC, Arnheim N, Warren ST (1997) The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J Med Genet* 34:627-631

Kusano K, Sunohara Y, Takahashi N, Yoshikura H, Kobayashi I (1994a) DNA double-strand break repair: genetic determinants of flanking crossing-over. *Proc Natl Acad Sci U S A* 91:1173-1177

Kusano K, Takahashi NK, Yoshikura H, Kobayashi I (1994b) Involvement of RecE exonuclease and RecT annealing protein in DNA double-strand break repair by homologous recombination. *Gene* 138:17-25

Kuzminov A (1995a) Instability of inhibited replication forks in E. coli. *Bioessays* 17:733-741

Kuzminov A (1995b) Collapse and repair of replication forks in Escherichia coli. *Mol Microbiol* 16:373-384

Kuzminov A (1996) Unraveling the late stages of recombinational repair: metabolism of DNA junctions in Escherichia coli. *Bioessays* 18:757-765

Kuzminov A (1999) Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* 63:751-813, table

Laban A, Cohen A (1981) Interplasmidic and intraplasmidic recombination in Escherichia coli K-12. *Mol Gen Genet* 184:200-207

Leach D, Lindsey J (1986) In vivo loss of supercoiled DNA carrying a palindromic sequence. *Mol Gen Genet* 204:322-327

Leach DR, Okely EA, Pinder DJ (1997) Repair by recombination of DNA containing a palindromic sequence. *Mol Microbiol* 26:597-606

Lee JS, Johnson DA, Morgan AR (1979) Complexes formed by (pyrimidine)n . (purine)n DNAs on lowering the pH are three-stranded. *Nucleic Acids Res* 6:3073-3091

Lee MK, Borchelt DR, Wong PC, Sisodia SS, Price DL (1996) Transgenic models of neurodegenerative diseases. *Curr Opin Neurobiol* 6:651-660

Lee MS, Marians KJ (1989) The Escherichia coli primosome can translocate actively in either direction along a DNA strand. *J Biol Chem* 264:14531-14542

Lee MS, Marians KJ (1990) Differential ATP requirements distinguish the DNA translocation and DNA unwinding activities of the Escherichia coli PRI A protein. *J Biol Chem* 265:17078-17083

Leeflang EP, Zhang L, Tavare S, Hubert R, Srinidhi J, MacDonald ME, Myers RH, de Young M, Wexler NS, Gusella JF (1995) Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum Mol Genet* 4:1519-1526

Leonhardt H, Lurz R, Alonso JC (1991) Physical and biochemical characterization of recombination-dependent synthesis of linear plasmid multimers in Bacillus subtilis. *Nucleic Acids Res* 19:497-503

Lewis LK, Harlow GR, L.A., Mount DW (1994) Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in Escherichia coli. *J Mol Biol* 241:507-523

Lilley DM (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci U S A* 77:6468-6472

Lilley DM, Markham AF (1983) Dynamics of cruciform extrusion in supercoiled DNA: use of a synthetic inverted repeat to study conformational populations. *EMBO* 2:527-533

Lilley DM (1985) The kinetic properties of cruciform extrusion are determined by DNA base-sequence. *Nucleic Acids Res* 13:1443-1465

Lin CT, Lyu YL, Liu LF (1997) A cruciform-dumbbell model for inverted dimer formation mediated by inverted repeats. *Nucleic Acids Res* 25:3009-3016

Lin LS, Meyer RJ (1987) DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Res* 15:8319-8331

Lindsey JC, Leach DR (1989) Slow replication of palindrome-containing DNA. *J Mol Biol* 206:779-782

Liu J, Nurse P, Marians KJ (1996) The ordered assembly of the phiX174-type primosome. III. PriB facilitates complex formation between PriA and DnaT. *J Biol Chem* 271:15656-15661

Liu J, Marians KJ (1999) PriA-directed assembly of a primosome on D loop DNA. *J Biol Chem* 274:25033-25041

Liu J, Xu L, Sandler SJ, Marians KJ (1999) Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc Natl Acad Sci U S A* 96:3552-3555

Lloyd RG, Picksley SM, Prescott C (1983) Inducible expression of a gene specific to the RecF pathway for recombination in Escherichia coli K12. *Mol Gen Genet* 190:162-167

Lloyd RG, Benson FE, Shurvinton CE (1984) Effect of ruv mutations on recombination and DNA repair in Escherichia coli K12. *Mol Gen Genet* 194:303-309

Lloyd RG, Buckman C (1985) Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J Bacteriol* 164:836-844

Lloyd RG, Porton MC, Buckman C (1988) Effect of recF, recJ, recN, recO and ruv mutations on ultraviolet survival and genetic recombination in a recD strain of Escherichia coli K12. *Mol Gen Genet* 212:317-324

Lloyd RG (1991) Conjugational recombination in resolvase-deficient ruvC mutants of Escherichia coli K-12 depends on recG. *J Bacteriol* 173:5414-5418

Lloyd RG, Buckman C (1991) Overlapping functions of recD, recJ and recN provide evidence of three epistatic groups of genes in Escherichia coli recombination and DNA repair. *Biochimie* 73:313-320

Lloyd RG, Sharples GJ (1993) Processing of recombination intermediates by the RecG and RuvAB proteins of Escherichia coli. *Nucleic Acids Res* 21:1719-1725

Loeb LA, Johansson E, Burgers PM, Fry M (2001) Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J Biol Chem* 276:16439-16446

Lovett ST, Luisi-DeLuca C, Kolodner RD (1988) The genetic dependence of recombination in recD mutants of Escherichia coli. *Genetics* 120:37-45

Lovett ST, Kolodner RD (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of Escherichia coli. *Proc Natl Acad Sci U S A* 86:2627-2631

Lovett ST, Drapkin PT, Sutera VA, Gluckman P (1993) A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in Escherichia coli. *Genetics* 135:631-642

Lovett ST, Feschenko VV (1996) Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. *Proc Natl Acad Sci U S A* 93:7120-7124

Luisi-DeLuca C, Lovett ST, Kolodner RD (1989) Genetic and physical analysis of plasmid recombination in recB recC sbcB and recB recC sbcA Escherichia coli K-12 mutants. *Genetics* 122:269-278

Lusetti SL, Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* 71:71-100

Lusky M, Hobom G (1979) Inceptor and origin of DNA replication in lambdoid coliphages. I. The lambda DNA minimal replication system. *Gene* 6:137-172

Lyons-Darden T, Topal MD (1999) Effects of temperature, Mg2+ concentration and mismatches on triplet-repeat expansion during DNA replication in vitro. *Nucleic Acids Res* 27:2235-2240

Lyu YL, Lin CT, Liu LF (1999) Inversion/dimerization of plasmids mediated by inverted repeats. *J Mol Biol* 285:1485-1501

MacFarland KJ, Shan Q, Inman RB, Cox MM (1997) RecA as a motor protein. Testing models for the role of ATP hydrolysis in DNA strand exchange. *J Biol Chem* 272:17675-17685

Madiraju MV, Clark AJ (1991) Effect of RecF protein on reactions catalyzed by RecA protein. *Nucleic Acids Res* 19:6295-6300

Magee TR, Asai T, Malka D, Kogoma T (1992) DNA damage-inducible origins of DNA replication in Escherichia coli. *EMBO* 11:4219-4225

Mahdi AA, Lloyd RG (1989) Identification of the recR locus of Escherichia coli K-12 and analysis of its role in recombination and DNA repair. *Mol Gen Genet* 216:503-510

Maisnier-Patin S, Nordstrom K, Dasgupta S (2001) RecA-mediated rescue of Escherichia coli strains with replication forks arrested at the terminus. *J Bacteriol* 183:6065-6073]

Marians KJ (1992) Prokaryotic DNA replication. *Annu Rev Biochem* 61:673-719

Marians KJ (1999) PriA: at the crossroads of DNA replication and recombination. *Prog Nucleic Acid Res Mol Biol* 63:39-67

Maniatis et al 1988

Masai H, Asai T, Kubota Y, Arai K, Kogoma T (1994) Escherichia coli PriA protein is essential for inducible and constitutive stable DNA replication. *EMBO J* 13:5338-5345

Maurer DJ, O'Callaghan BL, Livingston DM (1996) Orientation dependence of trinucleotide CAG repeat instability in Saccharomyces cerevisiae. *Mol Cell Biol* 16:6617-6622

Maurer DJ, O'Callaghan BL, Livingston DM (1998) Mapping the polarity of changes that occur in interrupted CAG repeat tracts in yeast. *Mol Cell Biol* 18:4597-4604

Mazin AV, Kuzminov AV, Dianov GL, Salganik RI (1991) Mechanisms of deletion formation in Escherichia coli plasmids. II. Deletions mediated by short direct repeats. *Mol Gen Genet* 228:209-214

McDaniel LD, Schultz RA (1992) Elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15. *Proc Natl Acad Sci U S A* 89:7968-7972

McEntee K, Weinstock GM, Lehman IR (1979) Initiation of general recombination catalyzed in vitro by the recA protein of Escherichia coli. *Proc Natl Acad Sci U S A* 76:2615-2619

McEntee K, Weinstock GM, Lehman IR (1981) Binding of the recA protein of Escherichia coli to single- and double-stranded DNA. *J Biol Chem* 256:8835-8844

McGlynn P, Al-Deib AA, Liu J, Marians KJ, Lloyd RG (1997) The DNA replication protein PriA and the recombination protein RecG bind D-loops. *J Mol Biol* 270:212-221

McGlynn P, Lloyd RG (1999) RecG helicase activity at three- and four-strand DNA structures. *Nucleic Acids Res* 27:3049-3056

McGlynn P, Lloyd RG (2000) Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* 101:35-45

McGlynn P, Mahdi AA, Lloyd RG (2000) Characterisation of the catalytically active form of RecG helicase. *Nucleic Acids Res* 28:2324-2332

McGlynn P, Lloyd RG (2000) Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* 101:35-45

McGlynn P, Lloyd RG (2001) Action of RuvAB at replication fork structures. *J Biol Chem* 276:41938-41944

McGlynn P, Lloyd RG, Marians KJ (2001) Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA is negatively supercoiled. *Proc Natl Acad Sci U S A* 98:8235-8240

McGlynn P, Lloyd RG (2002) Replicating past lesions in DNA. *Mol Cell* 10:700-701

McGlynn P, Lloyd RG (2002) Recombinational repair and restart of damaged replication forks. *Nat Rev Mol Cell Biol* 3:859-870

Menetski JP, Bear DG, Kowalczykowski SC (1990) Stable DNA heteroduplex formation catalyzed by the Escherichia coli RecA protein in the absence of ATP hydrolysis. *Proc Natl Acad Sci U S A* 87:21-25

Menetski JP, Kowalczykowski SC (1990) Biochemical properties of the Escherichia coli recA430 protein. Analysis of a mutation that affects the interaction of the ATP-recA protein complex with single-stranded DNA. *J Mol Biol* 211:845-855

Meyne J, Goodwin EH, Moyzis RK (1994) Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. *Chromosoma* 103:99-103

Michel B, Ehrlich SD, Uzest M (1997) DNA double-strand breaks caused by replication arrest. *EMBO* 16:430-438

Michel B, Recchia GD, Penel C, Ehrlich SD, Sherratt DJ (2000) Resolution of holliday junctions by RuvABC prevents dimer formation in rep mutants and UV-irradiated cells. *Mol Microbiol* 37:180-191

Michel B, Recchia GD, Penel-Colin M, Ehrlich SD, Sherratt DJ (2000) Resolution of holliday junctions by RuvABC prevents dimer formation in rep mutants and UV-irradiated cells. *Mol Microbiol* 37:180-191

Michel B (2000) Replication fork arrest and DNA recombination. *Trends Biochem Sci* 25:173-178

Miguel AG, Tyrrell RM (1986) Repair of near-ultraviolet (365 nm)-induced strand breaks in Escherichia coli DNA. The role of the polA and recA gene products. *Biophys J* 49:485-491

Miller J H (1992) A short course in bacterial genetics. A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbour Laboratory Press.

Miret JJ, Pessoa-Brandao L, Lahue RS (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 95:12438-12443

Mitas M, Yu A, Dill J, Kamp TJ, Chambers EJ, Haworth IS (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res* 23:1050-1059

Mitas M, Yu A, Dill J, Haworth IS (1995) The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry* 34:12803-12811

Mitas M (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res* 25:2245-2254

Mitchell AH, West SC (1996) Role of RuvA in branch migration reactions catalyzed by the RuvA and RuvB proteins of Escherichia coli. *J Biol Chem* 271:19497-19502

Mizuuchi K, Mizuuchi M, Gellert M (1982) Cruciform structures in palindromic DNA are favored by DNA supercoiling. *J Mol Biol* 156:229-243

Modrich P (1994) Mismatch repair, genetic stability, and cancer. *Science* 266:1959-1960

Modrich P (1997) Strand-specific mismatch repair in mammalian cells. *J Biol Chem* 272:24727-24730

Moore H, Greenwell PW, Liu CP, Arnheim N, Petes TD (1999) Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc Natl Acad Sci U S A* 96:1504-1509

Morag AS, Saveson CJ, Lovett ST (1999) Expansion of DNA repeats in Escherichia coli: effects of recombination and replication functions. *J Mol Biol* 289:21-27

Morel P, Reverdy C, Michel B, Ehrlich SD, Cassuto E (1998) The role of SOS and flap processing in microsatellite instability in Escherichia coli. *Proc Natl Acad Sci U S A* 95:10003-10008

Moser MJ, Oshima J, Monnat RJJ (1999) WRN mutations in Werner syndrome. *Hum Mutat* 13:271-279

Mosig G (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu Rev Genet* 32:379-413

Motamedi MR, Szigety SK, Rosenberg SM (1999) Double-strand-break repair recombination in Escherichia coli: physical evidence for a DNA replication mechanism in vivo. *Genes Dev* 13:2889-2903

Mount DW, Walker AC, Kosel C (1975) Indirect suppression of radiation sensitivity of a recA- strain of Escherichia coli K12. *Basic Life Sci* 5A:383-388

Mount DW (1977) A mutant of Escherichia coli showing constitutive expression of the lysogenic induction and error-prone DNA repair pathways. *Proc Natl Acad Sci U S A* 74:300-304

Muller B, Tsaneva IR, West SC (1993) Branch migration of Holliday junctions promoted by the Escherichia coli RuvA and RuvB proteins. II. Interaction of RuvB with DNA. *J Biol Chem* 268:17185-17189

Muller B, Tsaneva IR, West SC (1993) Branch migration of Holliday junctions promoted by the Escherichia coli RuvA and RuvB proteins. I. Comparison of RuvAB- and RuvB-mediated reactions. *J Biol Chem* 268:17179-17184

Murchie AI, Lilley DM (1987) The mechanism of cruciform formation in supercoiled DNA: initial opening of central basepairs in salt-dependent extrusion. *Nucleic Acids Res* 15:9641-9654

Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* 180:2063-2071

Murray J, Buard J, Neil DL, Yeramian E, Tamaki K, Hollies C, Jeffreys AJ (1999) Comparative sequence analysis of human minisatellites showing meiotic repeat instability. *Genome Res* 9:130-136

Murray JM, Lindsay HD, Munday CA, Carr AM (1997) Role of Schizosaccharomyces pombe RecQ homolog, recombination, and checkpoint genes in UV damage tolerance. *Mol Cell Biol* 17:6868-6875

Muyrers JP, Zhang Y, Buchholz F, Stewart AF (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* 14:1971-1982

Myung K, Datta A, Chen C, Kolodner RD (2001) SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination. *Nat Genet* 27:113-116

Nadel Y, Weisman-Shomer P, Fry M (1995) The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J Biol Chem* 270:28970-28977

Nag DK, Kurst A (1997) A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast Saccharomyces cerevisiae. *Genetics* 146:835-847

Nakayama H, Nakayama K, Nakayama R, Irino N, Nakayama Y, Hanawalt PC (1984) Isolation and genetic characterization of a thymineless death-resistant mutant of Escherichia coli K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway. *Mol Gen Genet* 195:474-480

Nakayama K, Irino N, Nakayama H (1985) The recQ gene of Escherichia coli K12: molecular cloning and isolation of insertion mutants. *Mol Gen Genet* 200:266-271

Nakayama K, Shiota S, Nakayama H (1988) Thymineless death in Escherichia coli mutants deficient in the RecF recombination pathway. *Can J Microbiol* 34:905-907

Naom IS, Morton SJ, Leach DR, Lloyd RG (1989) Molecular organization of sbcC, a gene that affects genetic recombination and the viability of DNA palindromes in Escherichia coli K-12. *Nucleic Acids Res* 17:8033-8045

Napierala M, Parniewski P, Pluciennik A, Wells RD (2002) Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination. *J Biol Chem* 277:34087-34100

Nasar F, Jankowski C, Nag DK (2000) Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol Cell Biol* 20:3449-3458

Newman BJ, Masters M (1980) The variation in frequency with which markers are transduced by phage P1 is primarily a result of discrimination during recombination. *Mol Gen Genet* 180:585-589

Noirot P, Bargonetti J, Novick RP (1990) Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc Natl Acad Sci U S A* 87:8560-8564

Noirot P, Kolodner RD (1998) DNA strand invasion promoted by Escherichia coli RecT protein. *J Biol Chem* 273:12274-12280

Nurse P, DiGate RJ, Zavitz KH, Marians KJ (1990) Molecular cloning and DNA sequence analysis of Escherichia coli priA, the gene encoding the primosomal protein replication factor Y. *Proc Natl Acad Sci U S A* 87:4615-4619

Nurse P, Zavitz KH, Marians KJ (1991) Inactivation of the Escherichia coli priA DNA replication protein induces the SOS response. *J Bacteriol* 173:6686-6693

Nurse P, Liu J, Marians KJ (1999) Two modes of PriA binding to DNA. *J Biol Chem* 274:25026-25032

Ohshima K, Kang S, Larson JE, Wells RD (1996) Cloning, characterization, and properties of seven triplet repeat DNA sequences. *J Biol Chem* 271:16773-16783

Ohshima K, Wells RD (1997) Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. *J Biol Chem* 272:16798-16806

Ohshima K, Montermini L, Wells RD, Pandolfo M (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. *J Biol Chem* 273:14588-14595

Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A* 59:598-605

Opperman T, Murli S, Smith BT, Walker GC (1999) A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc Natl Acad Sci U S A* 96:9218-9223

Ordway JM, Tallaksen-Greene S, Gutekunst CA, Bernstein EM, Cearley JA, Wiener HW, Dure LS, Lindsey R, Hersch SM, Jope RS, Albin RL, Detloff PJ (1997) Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* 91:753-763

Otsuji N, Iyehara H, Hideshima Y (1974) Isolation and characterization of an Escherichia coli ruv mutant which forms nonseptate filaments after low doses of ultraviolet light irradiation. *J Bacteriol* 117:337-344

Oussatcheva EA, Hashem VI, Zou Y, Sinden RR, Potaman VN (2001) Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in Escherichia coli. *J Biol Chem* 276:30878-30884

Pan X, Leach D R F (2000) The roles of *mutS*, *sbcCD*, and *recA* in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res* 28:3178-3184

Panayotatos N, Wells RD (1981) Cruciform structures in supercoiled DNA. *Nature* 289:466-470

Panigrahi GB, Cleary JD, Pearson CE (2002) In vitro (CTG)*(CAG) expansions and deletions by human cell extracts. *J Biol Chem* 277:13926-13934

Panyutin IG, Hsieh P (1993) Formation of a single base mismatch impedes spontaneous DNA branch migration. *J Mol Biol* 230:413-424

Paques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* 63:349-404

Parniewski P, Bacolla A, Jaworski A, Wells RD (1999) Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli. *Nucleic Acids Res* 27:616-623

Parsons CA, West SC (1988) Resolution of model Holliday junctions by yeast endonuclease is dependent upon homologous DNA sequences. *Cell* 52:621-629

Parsons CA, Tsaneva I, Lloyd RG, West SC (1992) Interaction of Escherichia coli RuvA and RuvB proteins with synthetic Holliday junctions. *Proc Natl Acad Sci U S A* 89:5452-5456

Parsons CA, Stasiak A, West SC (1995) The E.coli RuvAB proteins branch migrate Holliday junctions through heterologous DNA sequences in a reaction facilitated by SSB. *EMBO* 14:5736-5744

Paulson HL, Fischbeck KH (1996) Trinucleotide repeats in neurogenetic disorders. *Annu Rev Neurosci* 19:79-107

Pearson CE, Sinden RR (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* 35:5041-5053

Pearson CE, Ewel A, Acharya S, Fishel RA, Sinden RR (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet* 6:1117-1123

Pearson CE, Wang YH, Griffith JD, Sinden RR (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n. (CAG)n repeats from the myotonic dystrophy locus. *Nucleic Acids Res* 26:816-823

Pearson CE, Sinden RR (1998) Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr Opin Struct Biol* 8:321-330

Pearson CE, Eichler EE, Lorenzetti D, Kramer SF, Zoghbi HY, Nelson DL, Sinden RR (1998) Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* 37:2701-2708

Pearson CE, Wang YH, Griffith JD, Sinden RR (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n. (CAG)n repeats from the myotonic dystrophy locus. *Nucleic Acids Res* 26:816-823

Pearson CE, Tam M, Wang YH, Montgomery SE, Dar AC, Cleary JD, Nichol K (2002) Slipped-strand DNAs formed by long (CAG)*(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res* 30:4534-4547

Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 91:5355-5358

Petruska J, Arnheim N, Goodman MF (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res* 24:1992-1998

Petruska J, Hartenstine MJ, Goodman MF (1998) Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J Biol Chem* 273:5204-5210

Pham P, Seitz EM, Saveliev S, Shen X, Woodgate R, Cox MM, Goodman MF (2002) Two distinct modes of RecA action are required for DNA polymerase V-catalyzed translesion synthesis. *Proc Natl Acad Sci U S A* 99:11061-11066

Picksley SM, Attfield PV, Lloyd RG (1984) Repair of DNA double-strand breaks in Escherichia coli K12 requires a functional recN product. *Mol Gen Genet* 195:267-274

Picksley SM, Morton SJ, Lloyd RG (1985) The recN locus of Escherichia coli K12: molecular analysis and identification of the gene product. *Mol Gen Genet* 201:301-307

Pinder DJ, Blake CE, Leach DR (1997) DIR: a novel DNA rearrangement associated with inverted repeats. *Nucleic Acids Res* 25:523-529

Pinder DJ, Blake CE, Lindsey JC, Leach DR (1998) Replication strand preference for deletions associated with DNA palindromes. *Mol Microbiol* 28:719-727

Platt T (1986) Transcription termination and the regulation of gene expression. *Annu Rev Biochem* 55:339-372

Pluciennik A, Iyer RR, Napierala M, Larson JE, Filutowicz M, Wells RD (2002) Long CTG.CAG repeats from myotonic dystrophy are preferred sites for intermolecular recombination. *J Biol Chem* 277:34074-34086

Poteete AR, Fenton AC (2000) Genetic requirements of phage lambda red-mediated gene replacement in Escherichia coli K-12. *J Bacteriol* 182:2336-2340

Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM, Sauer RT (1980) How the lambda repressor and cro work. *Cell* 19:1-11

Radding CM, Shibata T, DasGupta C, Cunningham RP, Osber L (1981) Kinetics and topology of homologous pairing promoted by Escherichia coli recA-gene protein. *Cold Spring Harb Symp Quant Biol* 45 Pt 1:385-390

Rafferty JB, Sedelnikova SE, Hargreaves D, Artymiuk PJ, Baker PJ, Sharples GJ, Mahdi AA, Lloyd RG, Rice DW (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* 274:415-421

Rahmouni AR, Wells RD (1992) Direct evidence for the effect of transcription on local DNA supercoiling in vivo. *J Mol Biol* 223:131-144

Ramirez BE, Voloshin ON, R.D., Bax A (2000) Solution structure of DinI provides insight into its mode of RecA inactivation. *Protein Sci* 9:2161-2169

Ramstein J, Lavery R (1988) Energetic coupling between DNA bending and base pair opening. *Proc Natl Acad Sci U S A* 85:7231-7235

Ramstein J, Lavery R (1990) Base pair opening pathways in B-DNA. *J Biomol Struct Dyn* 7:915-933

Rangarajan S, Woodgate R, Goodman MF (1999) A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated Escherichia coli. *Proc Natl Acad Sci U S A* 96:9224-9229

Rangarajan S, Woodgate R, Goodman MF (2002) Replication restart in UV-irradiated Escherichia coli involving pols II, III, V, PriA, RecA and RecFOR proteins. *Mol Microbiol* 43:617-628

Register JC3, Griffith J (1985) The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. *J Biol Chem* 260:12308-12312

Rehrauer WM, Kowalczykowski SC (1996) The DNA binding site(s) of the Escherichia coli RecA protein. *J Biol Chem* 271:11996-12002

Rehrauer WM, Bruck I, Woodgate R, Goodman MF, Kowalczykowski SC (1998) Modulation of RecA nucleoprotein function by the mutagenic UmuD'C protein complex. *J Biol Chem* 273:32384-32387

Richard GF, Dujon B, Haber JE (1999) Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol Gen Genet* 261:871-882

Richard GF, Paques F (2000) Mini- and microsatellite expansions: the recombination connection. *EMBO Rep* 1:122-126

Richard GF, Goellner GM, McMurray CT, Haber JE (2000) Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. *EMBO* 19:2381-2390

Richards RI, Sutherland GR (1992) Heritable unstable DNA sequences. *Nat Genet* 1:7-9

Ricke DO, Liu Q, Gostout B, Sommer SS (1995) Nonrandom patterns of simple and cryptic triplet repeats in coding and noncoding sequences. *Genomics* 26:510-520

Riera J, de AR, Garriga X, Tapias A, Barbe J (1994) Interspecies regulation of the recA gene of gram-negative bacteria lacking an E. coli-like SOS operator. *Mol Gen Genet* 245:523-527

Robu ME, Inman RB, Cox MM (2001) RecA protein promotes the regression of stalled replication forks in vitro. *Proc Natl Acad Sci U S A* 98:8211-8218

Roman LJ, Dixon DA, Kowalczykowski SC (1991) RecBCD-dependent joint molecule formation promoted by the Escherichia coli RecA and SSB proteins. *Proc Natl Acad Sci U S A* 88:3367-3371

Rosche WA, Trinh TQ, Sinden RR (1995) Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J Bacteriol* 177:4385-4391

Rosche WA, Jaworski A, Kang S, Kramer SF, Larson JE, Geidroc DP, Wells RD, Sinden RR (1996) Single-stranded DNA-binding protein enhances the stability of CTG triplet repeats in Escherichia coli. *J Bacteriol* 178:5042-5044

Rothman RH, Clark AJ (1977) The dependence of postreplication repair on uvrB in a recF mutant of Escherichia coli K-12. *Mol Gen Genet* 155:279-286

Rothstein R, Michel B, Gangloff S (2000) Replication fork pausing and recombination or "gimme a break". *Genes Dev* 14:1-10

Rould E, Muniyappa K, Radding CM (1992) Unwinding of heterologous DNA by RecA protein during the search for homologous sequences. *J Mol Biol* 226:127-139

Ruskin B, Fink GR (1993) Mutations in POL1 increase the mitotic instability of tandem inverted repeats in Saccharomyces cerevisiae. *Genetics* 134:43-56

S.E., Lovett ST (1995) Enhancement of RecA strand-transfer activity by the RecJ exonuclease of Escherichia coli. *J Biol Chem* 270:6881-6885

Saito A, Iwasaki H, Ariyoshi M, Morikawa K, Shinagawa H (1995) Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase. *Proc Natl Acad Sci U S A* 92:7470-7474

Sakamoto N, Larson JE, Iyer RR, Montermini L, Pandolfo M, Wells RD (2001) GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. *J Biol Chem* 276:27178-27187

Sakamoto N, Ohshima K, Montermini L, Pandolfo M, Wells RD (2001) Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J Biol Chem* 276:27171-27177

Samadashwily GM, Raca G, Mirkin SM (1997) Trinucleotide repeats affect DNA replication in vivo. *Nat Genet* 17:298-304

Sambrook, Fritsch and Maniatis (1989) Molecular Cloning . A laboratory manual. 2nd edition. Cold Spring Harbour Laboratory Press.

Sancar A, Stachelek C, Konigsberg W, Rupp WD (1980) Sequences of the recA gene and protein. *Proc Natl Acad Sci U S A* 77:2611-2615

Sanchez I, Xu CJ, Juo P, Kakizaka A, Blenis J, Yuan J (1999) Caspase-8 is required for cell death induced by expanded polyglutamine repeats. *Neuron* 22:623-633

Sandler SJ, Clark AJ (1994) RecOR suppression of recF mutant phenotypes in Escherichia coli K-12. *J Bacteriol* 176:3661-3672

Sandler SJ, Samra HS, Clark AJ (1996) Differential suppression of priA2::kan phenotypes in Escherichia coli K-12 by mutations in priA, lexA, and dnaC. *Genetics* 143:5-13

Sandler SJ (1996) Overlapping functions for recF and priA in cell viability and UV-inducible SOS expression are distinguished by dnaC809 in Escherichia coli K-12. *Mol Microbiol* 19:871-880

Sargentini NJ, Smith KC (1986) Quantitation of the involvement of the recA, recB, recC, recF, recJ, recN, lexA, radA, radB, uvrD, and umuC genes in the repair of X-ray-induced DNA double-strand breaks in Escherichia coli. *Radiat Res* 107:58-72

Sarkar PS, Chang HC, Boudi FB, Reddy S (1998) CTG repeats show bimodal amplification in E. coli. *Cell* 95:531-540

Sassanfar M, Roberts JW (1990) Nature of the SOS-inducing signal in Escherichia coli. The involvement of DNA replication. *J Mol Biol* 212:79-96

Saudou F, Finkbeiner S, Devys D, Greenberg ME (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95:55-66

Sauer RT, Ross MJ, Ptashne M (1982) Cleavage of the lambda and P22 repressors by recA protein. *J Biol Chem* 257:4458-4462

Saveson CJ, Lovett ST (1997) Enhanced deletion formation by aberrant DNA replication in Escherichia coli. *Genetics* 146:457-470

Schaeffer F, Yeramian E, Lilley DM (1989) Long-range structural effects in supercoiled DNA: statistical thermodynamics reveals a correlation between calculated cooperative melting and contextual influence on cruciform extrusion. *Biopolymers* 28:1449-1473

Schmidt KH, Abbott CM, Leach DR (2000) Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli. *Mol Microbiol* 35:463-471

Schulz JB, Dichgans J (1999) Molecular pathogenesis of movement disorders: are protein aggregates a common link in neuronal degeneration? *Curr Opin Neurol* 12:433-439

Schumacher S, Fuchs RP, Bichara M (1998) Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in Escherichia coli: the effect of long patch mismatch repair revisited. *J Mol Biol* 279:1101-1110

Schumacher S, Pinet I, Bichara M (2001) Modulation of transcription reveals a new mechanism of triplet repeat instability in Escherichia coli. *J Mol Biol* 307:39-49

Schutte BC, Cox MM (1987) Homology-dependent changes in adenosine 5'-triphosphate hydrolysis during recA protein promoted DNA strand exchange: evidence for long paranemic complexes. *Biochemistry* 26:5616-5625

Schweitzer JK, Livingston DM (1997) Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Hum Mol Genet* 6:349-355

Seigneur M, Bidnenko V, Ehrlich SD, Michel B (1998) RuvAB acts at arrested replication forks. *Cell* 95:419-430

Seigneur M, Ehrlich SD, Michel B (1999) recD sbcB sbcD mutants are deficient in recombinational repair of UV lesions by RecBC. *J Bacteriol* 181:6220-6221

Seigneur M, Ehrlich SD, Michel B (2000) RuvABC-dependent double-strand breaks in dnaBts mutants require recA. *Mol Microbiol* 38:565-574

Sharma B, Hill TM (1995) Insertion of inverted Ter sites into the terminus region of the Escherichia coli chromosome delays completion of DNA replication and disrupts the cell cycle. *Mol Microbiol* 18:45-61

Sharma RC, Smith KC (1987) Role of DNA polymerase I in postreplication repair: a reexamination with Escherichia coli delta polA. *J Bacteriol* 169:4559-4564

Sharples GJ, Whitby MC, Ryder L, Lloyd RG (1994) A mutation in helicase motif III of E. coli RecG protein abolishes branch migration of Holliday junctions. *Nucleic Acids Res* 22:308-313

Sharples GJ, Leach DR (1995) Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Mol Microbiol* 17:1215-1217

Shen P, Huang HV (1986) Homologous recombination in *E. coli*: dependence on substrate length and homology. *Genetics* 112(3):441-57

Shen JC, Loeb LA, Fry M (1998) Werner syndrome protein. II. Characterization of the integral 3' --> 5' DNA exonuclease. *J Biol Chem* 273:34145-34150

Shen JC, Loeb LA (2000) Werner syndrome exonuclease catalyzes structure-dependent degradation of DNA. *Nucleic Acids Res* 28:3260-3268

Shiga Y, Sekine Y, Ohtsubo E (1999) Transposition of IS1 circles. *Genes Cells* 4:551-561

Shimizu M, Gellibolian R, Oostra BA, Wells RD (1996) Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J Mol Biol* 258:614-626

Shinagawa H, Makino K, Amemura M, Kimura S, Iwasaki H, Nakata A (1988) Structure and regulation of the Escherichia coli ruv operon involved in DNA repair and recombination. *J Bacteriol* 170:4322-4329

Shinagawa H, Iwasaki H, Kato T, Nakata A (1988) RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci U S A* 85:1806-1810

Shurvinton CE, Stahl MM, Stahl FW (1987) Large palindromes in the lambda phage genome are preserved in a rec+ host by inhibiting lambda DNA replication. *Proc Natl Acad Sci U S A* 84:1624-1628

Silberstein Z, Cohen A (1987) Synthesis of linear multimers of OriC and pBR322 derivatives in Escherichia coli K-12: role of recombination and replication functions. *J Bacteriol* 169:3131-3137

Simic D, Vukovic-Gacic B, Ajanovic A, Knezevic-Vukcevic J (1991) Activation of RecA protein in recombination-deficient strains of Escherichia coli following DNA-damaging treatments. *Mutat Res* 254:255-262

Sinden RR, Broyles SS, Pettijohn DE (1983) Perfect palindromic lac operator DNA sequence exists as a stable cruciform structure in supercoiled DNA in vitro but not in vivo. *Proc Natl Acad Sci U S A* 80:1797-1801

Sinden RR, Zheng GX, Brankamp RG, Allen KN (1991) On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. *Genetics* 129:991-1005

Sinden RR, Potaman VN, Oussatcheva EA, Pearson CE, Lyubchenko YL, Shlyakhtenko LS (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* 27:53-65

Singh L (1995) Biological significance of minisatellites. *Electrophoresis* 16:1586-1595

Smith KC, Sharma RC (1987) A model for the recA-dependent repair of excision gaps in UV-irradiated Escherichia coli. *Mutat Res* 183:1-9

Smith BT, Walker GC (1998) Mutagenesis and more: umuDC and the Escherichia coli SOS response. *Genetics* 148:1599-1610

Song L, Fujimoto BS, Wu PG, Thomas JC, Shibata JH, Schurr JM (1990) Evidence for allosteric transitions in secondary structure induced by superhelical stress. *J Mol Biol* 214:307-326

Spiro C, Pelletier R, Rolfsmeier ML, Dixon MJ, Lahue RS, Gupta G, Park MS, Chen X, Mariappan SV, McMurray CT (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol Cell* 4:1079-1085

Stasiak A, Di Capua E (1982) The helicity of DNA in complexes with recA protein. *Nature* 299:185-186

Stewart E, Chapman CR, Al-Khodairy F, Carr AM, Enoch T (1997) rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. *EMBO* 16:2682-2692

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb Symp Quant Biol* 31:77-84

Stukenberg PT, Turner J, O'Donnell M (1994) An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. *Cell* 78:877-887

Suen IS, Rhodes JN, Christy M, McEwen B, Gray DM, Mitas M (1999) Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochim Biophys Acta* 1444:14-24

Sugimoto K, Okazaki T, Okazaki R (1968) Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages. *Proc Natl Acad Sci U S A* 60:1356-1362

Summers D (1998) Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol Microbiol* 29:1137-1145

Summers DK, Beton CW, Withers HL (1993) Multicopy plasmid instability: the dimer catastrophe hypothesis. *Mol Microbiol* 8:1031-1038

Sun H, Karow JK, Hickson ID, Maizels N (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem* 273:27587-27592

Sun H, Bennett RJ, Maizels N (1999) The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Res* 27:1978-1984

Sutherland GR (1988) The role of nucleotides in human fragile site expression. *Mutat Res* 200:207-213

Sutton MD, Opperman T, Walker GC (1999) The Escherichia coli SOS mutagenesis proteins UmuD and UmuD' interact physically with the replicative DNA polymerase. *Proc Natl Acad Sci U S A* 96:12373-12378

Sutton MD, Farrow MF, Burton BM, Walker GC (2001) Genetic interactions between the Escherichia coli umuDC gene products and the beta processivity clamp of the replicative DNA polymerase. *J Bacteriol* 183:2897-2909

Taddei F, Matic I, Radman M (1995) cAMP-dependent SOS induction and mutagenesis in resting bacterial populations. *Proc Natl Acad Sci U S A* 92:11736-11740

Takeuchi F, Hanaoka F, Goto M, Yamada M, Miyamoto T (1982) Prolongation of S phase and whole cell cycle in Werner's syndrome fibroblasts. *Exp Gerontol* 17:473-480

Takeuchi F, Hanaoka F, Goto M, Akaoka I, Hori T, Yamada M, Miyamoto T (1982) Altered frequency of initiation sites of DNA replication in Werner's syndrome cells. *Hum Genet* 60:365-368

Taki K, Horiuchi T (1999) The SOS response is induced by replication fork blockage at a Ter site located on a pUC-derived plasmid: dependence on the distance between ori and Ter sites. *Mol Gen Genet* 262:302-309

Taylor AF, Smith GR (1992) RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. *Proc Natl Acad Sci U S A* 89:5226-5230

Terzaghi E, Okada Y, Streisinger G, Emrich J, Inouye M, Tsugita A (1966) Change of a sequence of amino acids in phage T4 lysozyme by acridine-induced mutations. *Proc Natl Acad Sci U S A* 56:500-507

Thomas BJ, Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* 56:619-630

Thompson BJ, Camien MN, Warner RC (1976) Kinetics of branch migration in double-stranded DNA. *Proc Natl Acad Sci U S A* 73:2299-2303

Thoms B, Wackernagel W (1998) Interaction of RecBCD enzyme with DNA at double-strand breaks produced in UV-irradiated Escherichia coli: requirement for DNA end processing. *J Bacteriol* 180:5639-5645

Tishkoff DX, Filosi N, Gaida GM, Kolodner RD (1997) A novel mutation avoidance mechanism dependent on S. cerevisiae RAD27 is distinct from DNA mismatch repair. *Cell* 88:253-263

Toman Z, Dambly-Chaudiere C, Tenenbaum L, Radman M (1985) A system for detection of genetic and epigenetic alterations in Escherichia coli induced by DNA-damaging agents. *J Mol Biol* 186:97-105

Trinh TQ, Sinden RR (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli. *Nature* 352:544-547

Tsaneva IR, Muller B, West SC (1993) RuvA and RuvB proteins of Escherichia coli exhibit DNA helicase activity in vitro. *Proc Natl Acad Sci U S A* 90:1315-1319

Tseng YC, Hung JL, Wang TC (1994) Involvement of RecF pathway recombination genes in postreplication repair in UV-irradiated Escherichia coli cells. *Mutat Res* 315:1-9

Umezu K, Nakayama K, Nakayama H (1990) Escherichia coli RecQ protein is a DNA helicase. *Proc Natl Acad Sci U S A* 87:5363-5367

Umezu K, Kolodner RD (1994) Protein interactions in genetic recombination in Escherichia coli. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. *J Biol Chem* 269:30005-30013

Usdin K, Woodford KJ (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res* 23:4202-4209

Usdin K (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res* 26:4078-4085

Uzest M, Ehrlich SD, Michel B (1995) Lethality of rep recB and rep recC double mutants of Escherichia coli. *Mol Microbiol* 17:1177-1188

van Gool AJ, Hajibagheri NM, Stasiak A, West SC (1999) Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of holliday junction resolution. *Genes Dev* 13:1861-1870

van G, Shah R, Mezard C, West SC (1998) Functional interactions between the holliday junction resolvase and the branch migration motor of Escherichia coli. *EMBO J* 17:1838-1845

van G, Hajibagheri NM, Stasiak A, West SC (1999) Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of holliday junction resolution. *Genes Dev* 13:1861-1870

Vincent SD, Mahdi AA, Lloyd RG (1996) The RecG branch migration protein of Escherichia coli dissociates R-loops. *J Mol Biol* 264:713-721

Viswanathan VK, Krcmarik K, Cianciotto NP (1999) Template secondary structure promotes polymerase jumping during PCR amplification. *Biotechniques* 27:508-511

Vogelstein B, Gillespie D (1979) Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci U S A* 76:615-619

Volkert MR, Hartke MA (1984) Suppression of Escherichia coli recF mutations by recA-linked srfA mutations. *J Bacteriol* 157:498-506

Voloshin ON, Ramirez BE, Bax A, R.D. (2001) A model for the abrogation of the SOS response by an SOS protein: a negatively charged helix in DinI mimics DNA in its interaction with RecA. *Genes Dev* 15:415-427

Wang TC, Smith KC (1988) Different effects of recJ and recN mutations on the postreplication repair of UV-damaged DNA in Escherichia coli K-12. *J Bacteriol* 170:2555-2559

Wang Y, Sauerbier W (1989) Flanking AT-rich sequences may lower the activation energy of cruciform extrusion in supercoiled DNA. *Biochem Biophys Res Commun* 158:423-431

Wang PZ, Projan SJ, Henriquez V, Novick RP (1993) Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element. *EMBO* 12:45-52

Wang YH, Amirhaeri S, Kang S, Wells RD, Griffith JD (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* 265:669-671

Wang WS, Seki M, Yamaoka T, Seki T, Tada S, Katada T, Fujimoto H, Enomoto T (1998) Cloning of two isoforms of mouse DNA helicase Q1/RecQL cDNA; alpha form is expressed ubiquitously and beta form specifically in the testis. *Biochim Biophys Acta* 1443:198-202

Watanabe T, Furuse C, Sakaizumi S (1968) Transduction of various R factors by phage P1 in Escherichia coli and by phage P22 in Salmonella typhimurium. *J Bacteriol* 96:1791-1795

Watt PM, Louis EJ, Borts RH, Hickson ID (1995) Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation. *Cell* 81:253-260

Webb BL, Cox MM, Inman RB (1997) Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. *Cell* 91:347-356

Webb BL, Cox MM, Inman RB (1999) ATP hydrolysis and DNA binding by the Escherichia coli RecF protein. *J Biol Chem* 274:15367-15374

Wells RD (1996) Molecular basis of genetic instability of triplet repeats. *J Biol Chem* 271:2875-2878

Wells RD, Parniewski P, Pluciennik A, Bacolla A, Gellibolian R, Jaworski A (1998) Small slipped register genetic instabilities in Escherichia coli in triplet repeat sequences associated with hereditary neurological diseases. *J Biol Chem* 273:19532-19541

West SC, Cassuto E, Mursalim J, Howard-Flanders P (1980) Recognition of duplex DNA containing single-stranded regions by recA protein. *Proc Natl Acad Sci U S A* 77:2569-2573

West SC (1996) The RuvABC proteins and Holliday junction processing in Escherichia coli. *J Bacteriol* 178:1237-1241

West SC (1996) The RuvABC proteins and Holliday junction processing in Escherichia coli. *J Bacteriol* 178:1237-1241

West SC (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* 31:213-244

Whitby MC, Ryder L, Lloyd RG (1993) Reverse branch migration of Holliday junctions by RecG protein: a new mechanism for resolution of intermediates in recombination and DNA repair. *Cell* 75:341-350

Whitby MC, Vincent SD, Lloyd RG (1994) Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. *EMBO J* 13:5220-5228

Whitby MC, Lloyd RG (1995) Altered SOS induction associated with mutations in recF, recO and recR. *Mol Gen Genet* 246:174-179

Whitby MC, Lloyd RG (1998) Targeting Holliday junctions by the RecG branch migration protein of Escherichia coli. *J Biol Chem* 273:19729-19739

Winans SC, Elledge SJ, Krueger JH, Walker GC (1985) Site-directed insertion and deletion mutagenesis with cloned fragments in Escherichia coli. *J Bacteriol* 161:1219-1221

Witkin EM, Kogoma T (1984) Involvement of the activated form of RecA protein in SOS mutagenesis and stable DNA replication in Escherichia coli. *Proc Natl Acad Sci U S A* 81:7539-7543

Worth LJ, Clark S, Radman M, Modrich P (1994) Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci U S A* 91:3238-3241

Wu L, Karow JK, Hickson ID (1999) Genetic recombination: Helicases and topoisomerases link up. *Curr Biol* 9:R518-R520

Wu X, Li J, Li X, Hsieh CL, Burgers PM, Lieber MR (1996) Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucleic Acids Res* 24:2036-2043

Yamagata K, Kato J, Shimamoto A, Goto M, Furuichi Y, Ikeda H (1998) Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast sgs1 mutant: implication for genomic instability in human diseases. *Proc Natl Acad Sci U S A* 95:8733-8738

Yanisch-Perron C, Vieira J, Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33(1):103-19.

Youngs DA, Van der SE, Smith KC (1974) Separate branches of the uvr gene-dependent excision repair process in ultraviolet-irradiated Escherichia coli K-12 cells; their dependence upon growth medium and the polA, recA, recB, and exrA genes. *J Bacteriol* 117:717-725

Yu A, Barron MD, Romero RM, Christy M, Gold B, Dai J, Gray DM, Haworth IS, Mitas M (1997) At physiological pH, d(CCG)15 forms a hairpin containing protonated cytosines and a distorted helix. *Biochemistry* 36:3687-3699

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* 97:5978-5983

Zaman MM, Boles TC (1996) Plasmid recombination by the RecBCD pathway of Escherichia coli. *J Bacteriol* 178:3840-3845

Zaman MM, Boles TC (1996) Plasmid recombination by the RecBCD pathway of Escherichia coli. *J Bacteriol* 178:3840-3845

Zavitz KH, Marians KJ (1992) ATPase-deficient mutants of the Escherichia coli DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. *J Biol Chem* 267:6933-6940

Zerbib D, Mezard C, George H, West SC (1998) Coordinated actions of RuvABC in Holliday junction processing. *J Mol Biol* 281:621-630

Zhang L, Leeflang EP, Yu J, Arnheim N (1994) Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nat Genet* 7:531-535

Zhang Y, Muyrers JP, Testa G, Stewart AF (2000) DNA cloning by homologous recombination in Escherichia coli. *Nat Biotechnol* 18:1314-1317

Zheng G, Ussery DW, Sinden RR (1991) Estimation of superhelical density in vivo from analysis of the level of cruciforms existing in living cells. *J Mol Biol* 221:122-129

Zheng M, Huang X, Smith GK, Yang X, Gao X (1996) Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J Mol Biol* 264:323-336